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

December 30, 2019

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 correlation of individual decisions within and across different activities, and provide an additional channel to identify peer effects. 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 asymptotic 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

*I thank the editor and two anonymous reviewers for helpful comments and suggestions. All remaining errors are my own.

†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. In a recent paper, Cohen-Cole et al. (2018) develop a simultaneous equations network model and suggest an estimation procedure by extending the generalized spatial 2SLS and 3SLS estimators proposed by Kelejian and Prucha (2004). Following the insight of Bramoullé et al. (2009), the identification strategy in Cohen-Cole et al. (2018) exploits exclusion restrictions from 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 Bonacich centrality (Bonacich, 1987) provides additional information to identify peer effects and can be used as an instrumental variable (IV) to improve estimation efficiency. As Cohen-Cole et al. (2018) and Liu (2014) focus on IV-based linear estimators, the corresponding identification strategy only utilizes linear moment conditions.

For single-equation spatial econometric models, quadratic moment conditions capturing spatial correlation of cross-sectional units are often used for identification when the model cannot be identified through linear moment conditions (see, e.g., Kelejian and Prucha, 1999; Lee, 2007). In this paper, we propose quadratic moment conditions based on the correlation of individual choices within and across equations for the identification of simultaneous equations network models. The idea of identifying peer effects by the correlation of individual choices 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 correlation of individual choices in different activities within a group, and thus does not rely on variation in group sizes.

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 asymptotic efficiency of the IV-based linear estimators proposed by Liu (2014). Compared to the quasi-maximum likelihood estimator proposed by Yang and Lee (2017) for the simultaneous equations spatial autoregressive model, the GMM estimator is computationally simple and remains tractable with group fixed effects.[1] Liu and Saraiva (2015, 2019) also consider GMM estimation of the simultaneous equations spatial autoregressive model. Liu and Saraiva (2015) focus on the special case with a triangular system of equations, and Liu and Saraiva (2019) propose a robust GMM estimator under heteroskedasticity of unknown form. Compared with Liu and Saraiva (2015, 2019), this paper emphasizes the group structure of network data and studies the identification and asymptotic properties of the GMM estimator in the presence of group fixed effects.

The rest of the paper is organized as follows. Sections 2 introduces the econometric model. The GMM estimator is described in Section 3, with its identification conditions and asymptotic properties studied in Sections 4 and 5 respectively. Section 6 provides Monte Carlo evidence on the finite sample performance of the proposed estimator. Section 7 briefly concludes. The proofs are collected in the appendix.

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

[1] For a network model with group fixed effects, if the adjacency matrix has non-constant row sums, the fixed effects cannot be eliminated by the within transformation (see Section 2) from the reduced form equations. Therefore, the quasi-maximum likelihood estimator needs to estimate the fixed effects together with other model parameters, and thus could suffer the "incidental parameter" problem (Neyman and Scott, 1948).

let $A^{(s)} = A + A'$, $\mathrm{vec}_D(A) = (a_{11}, \cdots, a_{nn})'$, 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 $\max_{i=1,\cdots,n} \sum_{j=1}^{n} |a_{ij}|$ (or $\max_{j=1,\cdots,n} \sum_{i=1}^{n} |a_{ij}|$) is bounded as $n \to \infty$. For an $n \times m$ matrix $B = [b_{ij}]$, the vectorization of B is denoted by $\mathrm{vec}(B) = (b_{11}, b_{21}, \cdots, b_{nm})'$,[2] and the Euclidean matrix norm of B is denoted by $||B|| = \sqrt{\mathrm{tr}(B'B)}$. Let $\mathrm{diag}\{B_j\}$ denote a "generalized" block diagonal matrix with a typical diagonal block being an $n_j \times m_j$ matrix B_j. Let I_n denote the $n \times n$ identity matrix with its kth column denoted by $i_{n,k}$. Let ι_n denote an $n \times 1$ vector of ones.

2 Econometric Model

Suppose a population of n individuals is partitioned into $\bar{g}$ non-overlapping groups, with n_g individuals in the gth group. Individuals in the same group interact in m activities through a network. The network topology of the gth group is captured by an $n_g \times n_g$ zero-diagonal adjacency matrix $W_{(g)}$. The (i, j)th element of $W_{(g)}$ is a known nonnegative constant representing the proximity of individuals i and j in the network.

For the gth group, the choices of n_g individuals in m activities are given by a system of m equations:

$$Y_{(g)} = Y_{(g)}\Phi_0 + W_{(g)}Y_{(g)}\Lambda_0 + X_{(g)}B_0 + W_{(g)}X_{(g)}\Gamma_0 + \iota_{n_g}\eta_{(g)} + U_{(g)}, \tag{2.1}$$

where $Y_{(g)}$ is an $n_g \times m$ matrix of observations on m endogenous variables, $X_{(g)}$ is an $n_g \times k_x$ matrix of observations on k_x exogenous variables, $U_{(g)}$ is an $n_g \times m$ matrix of disturbances, and $\eta_{(g)}$ is a $1 \times m$ vector of group fixed effects. $\Phi_0 = [\phi_{lk,0}], \Lambda_0 = [\lambda_{lk,0}], B_0$ and Γ_0 are, respectively, $m \times m, m \times m, k_x \times m$ and $k_x \times m$ matrices of true parameters in the data

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

generating process (DGP). As a normalization, $\phi_{kk,0} = 0$ for all k. We assume that each row of $U_{(g)}$ 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 is motivated by the best response function of a multiple-activity network game introduced by Liu (2014) and Cohen-Cole et al. (2018). In model (2.1), there are three types of endogenous effects, namely, *the simultaneity effect* (given by $\phi_{lk,0}$), wherein an individual's choice in a certain activity is affected by her own choices in related activities; *the within-activity peer effect* (given by $\lambda_{kk,0}$), wherein an individual's choice in a certain activity is affected by her peers' choices in the same activity; and *the cross-activity peer effect* (given by $\lambda_{lk,0}$, $l \neq k$), wherein an individual's choice in a certain activity is affected by her peers' choices in related activities. In addition to these endogenous effects, Γ_0 represents the *contextual effect*, wherein an individual's choice is affected by the exogenous characteristics of her peers, and $\eta_{(g)}$ 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).

Let $Y = [Y'_{(1)}, \cdots, Y'_{(\overline{g})}]'$, $X = [X'_{(1)}, \cdots, X'_{(\overline{g})}]'$, $U = [U'_{(1)}, \cdots, U'_{(\overline{g})}]'$, $W = \mathrm{diag}\{W_{(g)}\}$, and $L = \mathrm{diag}\{\iota_{n_g}\}$. Then, for all the $\overline{g}$ groups,

$$Y = Y\Phi_0 + \overline{Y}\Lambda_0 + XB_0 + \overline{X}\Gamma_0 + L\eta + U, \tag{2.2}$$

where $\overline{Y} = WY$, $\overline{X} = WX$, and $\eta = [\eta'_{(1)}, \cdots, \eta'_{(\overline{g})}]'$.

In general, the identification of simultaneous equations models requires exclusion restrictions. Let $Y_k, \overline{Y}_k, X_k$ and $\overline{X}_k$ denote the matrices containing columns of $Y, \overline{Y}, X$ and $\overline{X}$ that appear in the kth equation under some exclusion restrictions, and let $\phi_{k,0}, \lambda_{k,0}, \beta_{k,0}$ and $\gamma_{k,0}$ denote the corresponding vectors of true parameters. Then, the kth equation of model (2.2) is

$$y_k = Y_k\phi_{k,0} + \overline{Y}_k\lambda_{k,0} + X_k\beta_{k,0} + \overline{X}_k\gamma_{k,0} + L\eta_k + u_k, \tag{2.3}$$

where y_k, η_k and u_k are respectively the kth columns of Y, η and U. Let $Z_k = [Y_k, \overline{Y}_k, X_k, \overline{X}_k]$ and $\theta_{k,0} = (\phi'_{k,0}, \lambda'_{k,0}, \beta'_{k,0}, \gamma'_{k,0})'$. Equation (2.3) can be written more compactly as

$$y_k = Z_k\theta_{k,0} + L\eta_k + u_k. \tag{2.4}$$

We allow η to depend on W and X by treating η as a $\overline{g} \times m$ matrix of unknown parameters. When the number of network $\overline{g}$ is large, we may have the "incidental parameter" problem (Neyman and Scott, 1948). To avoid this problem, we eliminate the group fixed effects using the "deviation from group mean" projector $J = \mathrm{diag}\{J_{(g)}\}$ where $J_{(g)} = I_{n_g} - n_g^{-1}\iota_{n_g}\iota'_{n_g}$. This transformation is analogous to the within transformation for the fixed-effect panel data model. As $JL = 0$, the kth equation of the within-transformed model is

$$Jy_k = JZ_k\theta_{k,0} + Ju_k.$$

We maintain the following assumptions regarding the DGP. 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 Σ is an $m \times m$ nonsingular matrix. For some $\delta > 0$, $\mathrm{E}|u_{ik}u_{il}u_{is}u_{it}|^{1+\delta}$ 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 $\rho(\Lambda_0(I_m - \Phi_0)^{-1}) < 1/\rho(W)$.

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 5. $\theta_{k,0}$ is in the interior of a compact and convex parameter space for $k = 1, \cdots, m$.

Assumption 1-5 are from Kelejian and Prucha (2004) and Yang and Lee (2017). In particular, Assumption 2 imposes a restriction on the parameter space so that model (2.2) has a well defined reduced form. Assumption 3 limits the interdependence between individuals' choices to a tractable degree. If W is specified as a binary indicator matrix such that its (i, j)th element is one 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.

3 GMM Estimation

Let

$$u_k(\theta_k) = J(y_k - Z_k\theta_k). \tag{3.1}$$

Inspired by the GMM estimator proposed by Lee (2007) 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. For an $n \times q$ nonstochastic IV matrix Q, define the within-transformed IV matrix as $\ddot{Q} = JQ$. The linear moment functions are given by

$$h_{1,k}(\theta_k) = \ddot{Q}' u_k(\theta_k), \qquad \text{for } k = 1, \cdots, m.$$

For an $n \times n$ nonstochastic weighting matrix Ξ_r $(r = 1, \cdots, p)$, define the within-transformed weighting matrix as $\ddot{\Xi}_r = J\Xi_r J - \mathrm{tr}(J\Xi_r)J/\mathrm{tr}(J)$. The quadratic moment functions are given by

$$h_{2,kl}(\theta_k,\theta_l) = [\ddot{\Xi}_1' u_k(\theta_k), \cdots, \ddot{\Xi}_p' u_k(\theta_k)]' u_l(\theta_l), \qquad \text{for } k,l = 1,\cdots,m.$$

These moment conditions are valid because, at the true parameter value, $\mathrm{E}[h_{1,k}(\theta_{k,0})] = \mathrm{E}(\ddot{Q}' u_k) = 0$ and $\mathrm{E}[h_{2,kl}(\theta_{k,0},\theta_{l,0})] = \mathrm{E}[(\ddot{\Xi}_1' u_k, \cdots, \ddot{\Xi}_p' u_k)' u_l] = \sigma_{kl}[\mathrm{tr}(\ddot{\Xi}_1), \cdots, \mathrm{tr}(\ddot{\Xi}_p)]' = 0$.

Let

$$\begin{aligned} h_1(\theta) &= [h_{1,1}(\theta_1)', \cdots, h_{1,m}(\theta_m)']', \\ h_2(\theta) &= [h_{2,11}(\theta_1,\theta_1)', h_{2,12}(\theta_1,\theta_2)', \cdots, h_{2,mm}(\theta_m,\theta_m)']', \end{aligned}$$

and

$$h(\theta) = [h_1(\theta)', h_2(\theta)']', \tag{3.2}$$

where $\theta = (\theta_1', \cdots, \theta_m')'$. The GMM estimator for θ_0 is given by

$$\widehat{\theta}_{gmm} = \arg\min h(\theta)' \widehat{\Omega}^{-1} h(\theta), \tag{3.3}$$

where $n^{-1}\widehat{\Omega}$ is a $\sqrt{n}$-consistent estimator of $n^{-1}\Omega := n^{-1}\mathrm{Var}[h(\theta_0)]$.[3]

The (infeasible) optimal IV matrix for JZ_k is $F_k := \mathrm{E}(JZ_k)$. As $\mathrm{E}(Y) = \sum_{j=0}^{\infty} W^j [X, L] C_j$, where C_j is a coefficient matrix whose elements are functions of the elements of $\Phi_0, \Lambda_0, B_0, \Gamma_0$ and η (Kelejian and Prucha, 2004), the optimal IV matrix F_k can be expressed as a linear combination of the IVs in $\ddot{Q}_\infty = J[X, WX, W^2X, \cdots, WL, W^2L, \cdots]$. As shown in Liu and Lee (2010) and Liu (2014), $WL, W^2L, \cdots$ are the leading order terms of the Bonacich centrality (Bonacich, 1987). If $W_{(g)}$ has constant row sums (including the case that $W_{(g)}$ is

[3]For notational simplicity, we assume that the moment functions in $h(\theta)$ are non-redundant. Otherwise, one could define the GMM estimator based on $Dh(\theta)$, where D is a selector matrix that selects all non-redundant moment functions from $h(\theta)$.

row-normalized) for all g, then $WL, W^2L, \cdots$ are eliminated by the within transformation and $\ddot{Q}_\infty$ becomes $J[X, WX, W^2X, \cdots]$. If $W_{(g)}$ has non-constant row sums for some g, then $JWL, JW^2L, \cdots$ provide additional information to identify peer effects. Therefore, in the latter case, the identification condition is in general weaker (Liu, 2014).

To implement the GMM estimator, the researcher may choose an IV matrix $\ddot{Q}$ containing a subset of the linearly independent columns of $\ddot{Q}_\infty$.[4] As JWL has $\overline{g}$ columns, where $\overline{g}$ is the number of groups in the data, the number of IVs is proportional to the number of groups when JWL is included in $\ddot{Q}$. If the number of groups increases with n to infinity, so does the number of IVs. Therefore, we follow the many-instrument asymptotics (Bekker, 1994) and allow q, the number of IVs in $\ddot{Q}$, to go to infinity. Furthermore, we assume that the (infeasible) optimal IV matrix F_k can be approximated by a linear combination of the IVs in $\ddot{Q}$, with the approximation error diminishes as $q \to \infty$. This assumption is common in the many-instrument literature (see, e.g., Donald and Newey, 2001; Hansen et al., 2008).[5]

Assumption 6. $\ddot{Q} = JQ$, where Q is an $n \times q$ constant matrix with uniformly bounded elements. $\ddot{Q}$ has full column rank for n sufficiently large. For each q, there exists a constant matrix $C_{k,q}$ such that $n^{-1}||F_k - \ddot{Q}C_{k,q}||^2 \to 0$ as $q, n \to \infty$, for $k = 1, \cdots, m$.

The weighting matrices $\ddot{\Xi}_1, \cdots, \ddot{\Xi}_p$ in the quadratic moment functions are constructed from $\Xi_1, \cdots, \Xi_p$ satisfying the following regularity condition. We assume p is a fixed positive integer chosen together with Ξ_r's by the researcher. Possible candidates for Ξ_r include W, W^2, etc.[6]

[4] For the single-equation spatial autoregressive model, Liu and Lee (2013) derive the approximate mean square error (MSE) of the 2SLS estimator as the criterion to choose the optimal set of IVs. Although beyond the scope of this paper, it might be possible to derive the approximate MSE of the GMM estimator for the simultaneous-equation network model following Liu and Lee (2013).

[5] For a primitive condition for Assumption 6 to hold, see Liu and Lee (2010, 2013) for more discussions.

[6] For the single-equation spatial autoregressive model, Liu and Lee (2010) derive the optimal Ξ_r assuming

Assumption 7. $\ddot{\Xi}_r = J\Xi_r J - \mathrm{tr}(J\Xi_r)J/\mathrm{tr}(J)$, where Ξ_r is an $n \times n$ constant matrix with uniformly bounded row and column sums, for $r = 1, \cdots, p$.

4 Asymptotic Identification and Consistency

For the GMM estimator, θ_0 is asymptotically identified if $\mathrm{plim}_{n\to\infty} n^{-1}h(\theta)'\widehat{\Omega}^{-1}h(\theta)$ attains a unique minimum at θ_0. As $n^{-1}h(\theta)'\widehat{\Omega}^{-1}h(\theta) = n^{-1}h(\theta)'\Omega^{-1}h(\theta) + o_p(1)$ uniformly in θ (see the proof of Proposition 4.2), an asymptotically equivalent identification condition is that $\mathrm{plim}_{n\to\infty} n^{-1}h(\theta)'\Omega^{-1}h(\theta)$ attains a unique minimum at θ_0.

Let $\ddot{P} = \ddot{Q}(\ddot{Q}'\ddot{Q})^{-1}\ddot{Q}'$, $\omega = [\mathrm{vec}_D(\ddot{\Xi}_1), \cdots, \mathrm{vec}_D(\ddot{\Xi}_p)]$, and

$$\mu_3 = \begin{bmatrix} \mathrm{E}(u_{i1}u_{i1}u_{i1}) & \mathrm{E}(u_{i1}u_{i1}u_{i2}) & \cdots & \mathrm{E}(u_{i1}u_{im}u_{im}) \\ \vdots & \vdots & \ddots & \vdots \\ \mathrm{E}(u_{im}u_{i1}u_{i1}) & \mathrm{E}(u_{im}u_{i1}u_{i2}) & \cdots & \mathrm{E}(u_{im}u_{im}u_{im}) \end{bmatrix}.$$

Then,

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

where $\Omega_{11} = \mathrm{Var}[h_1(\theta_0)] = \Sigma \otimes (\ddot{Q}'\ddot{Q})$, $\Omega_{12} = \mathrm{E}[h_1(\theta_0)h_2(\theta_0)'] = \mu_3 \otimes (\ddot{Q}'\omega)$, and $\Omega_{22} = \mathrm{Var}[h_2(\theta_0)]$. The (infeasible) GMM objective function can be written as

$$h(\theta)'\Omega^{-1}h(\theta) = h_1(\theta)'\Omega_{11}^{-1}h_1(\theta) + h_2^*(\theta)'\Omega_{22}^{*-1}h_2^*(\theta) = u(\theta)'(\Sigma^{-1} \otimes \ddot{P})u(\theta) + h_2^*(\theta)'\Omega_{22}^{*-1}h_2^*(\theta),$$

the disturbances follow a normal distribution. For a specific simultaneous-equation network model (e.g., the model considered in Example 1), the optimal Ξ_r may exist under proper assumptions (e.g., normality of the disturbances).

where

$$
\begin{aligned}
h_2^*(\theta) &= h_2(\theta) - \Omega_{12}'\Omega_{11}^{-1}h_1(\theta) = h_2(\theta) - [(\mu_3'\Sigma^{-1}) \otimes (\omega'\ddot{P})]u(\theta), \\
\Omega_{22}^* &= \mathrm{Var}[h_2^*(\theta)] = \Omega_{22} - \Omega_{12}'\Omega_{11}^{-1}\Omega_{12} = \Omega_{22} - (\mu_3'\Sigma^{-1}\mu_3) \otimes (\omega'\ddot{P}\omega).
\end{aligned}
$$

Let $F = \mathrm{diag}\{F_k\}$, where $F_k = \mathrm{E}(JZ_k)$. As

$$
\begin{aligned}
n^{-1}u(\theta)'(\Sigma^{-1} \otimes \ddot{P})u(\theta) &= n^{-1}(\theta_0 - \theta)'F'(\Sigma^{-1} \otimes I_n)F(\theta_0 - \theta) + o_p(1) \\
n^{-1}h_2^*(\theta)'\Omega_{22}^{*-1}h_2^*(\theta) &= n^{-1}\overline{h}_2^*(\theta)'\Omega_{22}^{*-1}\overline{h}_2^*(\theta) + o_p(1)
\end{aligned}
$$

where $\overline{h}_2^*(\theta) = \mathrm{E}[h_2(\theta)] - [(\mu_3'\Sigma^{-1}) \otimes \omega']F(\theta_0 - \theta)$, uniformly in θ (see the proof of Proposition 4.2), θ_0 is asymptotically identified if

$$
\mathrm{lim}_{n\to\infty} n^{-1}(\theta_0 - \theta)'F'(\Sigma^{-1} \otimes I_n)F(\theta_0 - \theta) + \lim_{n\to\infty} n^{-1}\overline{h}_2^*(\theta)'\Omega_{22}^{*-1}\overline{h}_2^*(\theta)
$$

attains a unique minimum at θ_0. We assume the following regularity conditions. Let F_k^* be a matrix containing all the linearly independent columns of F_k and $F^* = \mathrm{diag}\{F_k^*\}$.

Assumption 8. (i) $\lim_{n\to\infty} n^{-1}F^{*\prime}(\Sigma^{-1} \otimes I_n)F^*$ is a finite and nonsingular matrix, (ii) $\lim_{n\to\infty} n^{-1}[(\mu_3'\Sigma^{-1}) \otimes \omega']F^*$ is a finite matrix with full column rank, and (iii) $\lim_{n\to\infty} n^{-1}\Omega_{22}^*$ is a finite and nonsingular matrix.

As $(\theta_0 - \theta)'F'(\Sigma^{-1} \otimes I_n)F(\theta_0 - \theta) \geq 0$ and $\overline{h}_2^*(\theta)'\Omega_{22}^{*-1}\overline{h}_2^*(\theta) \geq \overline{h}_2^*(\theta_0)'\Omega_{22}^{*-1}\overline{h}_2^*(\theta_0) = 0$, the asymptotic identification fails only if

$$
\mathrm{lim}_{n\to\infty} n^{-1}(\theta_0 - \theta)'F'(\Sigma^{-1} \otimes I_n)F(\theta_0 - \theta) = 0 \tag{4.1}
$$

and

$$\lim_{n\to\infty} n^{-1}\overline{h}_2^*(\theta)'\Omega_{22}^{*-1}\overline{h}_2^*(\theta) = 0 \tag{4.2}$$

for some $\theta \neq \theta_0$. If $\lim_{n\to\infty} n^{-1}F'(\Sigma^{-1}\otimes I_n)F$ is nonsingular, then θ_0 is asymptotically identified from (4.1). If $\lim_{n\to\infty} n^{-1}F'(\Sigma^{-1}\otimes I_n)F$ is singular, then (4.2) provides an additional channel for asymptotic identification. More specifically, suppose $F(\theta_0-\theta) = F^*(\theta_0^{(1)}-\theta^{(1)}) + F^*C(\theta_0^{(2)}-\theta^{(2)})$ for some constant matrix C, where $\theta_0^{(1)}$ is a vector of coefficients corresponding to the linearly independent columns of F. Then, under Assumption 8, the solutions of (4.1) are characterized by $\theta^{(1)} = \theta_0^{(1)} + C(\theta_0^{(2)}-\theta^{(2)})$. Thus, $\theta_0^{(1)}$ is identified if $\theta_0^{(2)}$ can be identified. Substitution of $\theta^{(1)} = \theta_0^{(1)} + C(\theta_0^{(2)}-\theta^{(2)})$ into (4.2), we can show that (4.2) has a unique solution at θ_0 if and only if

$$\lim_{n\to\infty} n^{-1}\mathrm{E}[h_2(\theta)] = 0, \qquad \text{s.t. } \theta^{(1)} = \theta_0^{(1)} + C(\theta_0^{(2)}-\theta^{(2)}), \tag{4.3}$$

has a unique solution at $\theta_0^{(2)}$. Therefore, asymptotic identification is achieved if $\theta_0^{(2)}$ can be identified from (4.3).

In the following, we discuss the asymptotic identification in more detail following the two-step identification strategy in Yang and Lee (2017), where the "pseudo" reduced form parameters is identified in the first step, and the structural parameters are recovered from the "pseudo" reduced form parameters in the second step.

Identification of the pseudo reduced form parameters Model (2.2) has a "pseudo" reduced form

$$Y = \overline{Y}\Lambda_0^* + XB_0^* + \overline{X}\Gamma_0^* + L\eta^* + U^*, \tag{4.4}$$

where

$$\Lambda_0^* = \Lambda_0(I_m-\Phi_0)^{-1}, \quad B_0^* = B_0(I_m-\Phi_0)^{-1}, \quad \Gamma_0^* = \Gamma_0(I_m-\Phi_0)^{-1}, \tag{4.5}$$

$\eta^* = \eta(I_m - \Phi_0)^{-1}$ and $U^* = U(I_m - \Phi_0)^{-1}$. The kth equation of the "pseudo" reduced form is given by

$$y_k = \sum_{l=1}^m \lambda^*_{lk,0}\overline{y}_l + X\beta^*_{k,0} + \overline{X}\gamma^*_{k,0} + L\eta^*_k + u^*_k, \qquad (4.6)$$

with

$$\overline{y}_k = G_k[(B_0^{*\prime} \otimes I_n + \Gamma_0^{*\prime} \otimes W)\text{vec}(X) + (I_m \otimes L)\text{vec}(\eta^*) + \text{vec}(U^*)], \qquad (4.7)$$

where $G_k = (i'_{m,k} \otimes W)(I_{mn} - \Lambda_0^{*\prime} \otimes W)^{-1}$. As $(u^*_{i1}, \cdots, u^*_{im})' \sim i.i.d.(0, \Sigma^*)$, where $\Sigma^* = (I_m - \Phi_0')^{-1}\Sigma(I_m - \Phi_0)^{-1}$, the "pseudo" reduced form parameters can be estimated by the GMM estimator defined in (3.3) with the linear and quadratic moment functions given by

$$\begin{aligned} h_{1,k}(\theta^*_k) &= \ddot{Q}' u^*_k(\theta^*_k), \\ h_{2,kl}(\theta^*_k, \theta^*_l) &= [\ddot{\Xi}_1' u^*_k(\theta^*_k), \cdots, \ddot{\Xi}_p' u^*_k(\theta^*_k)]' u^*_l(\theta^*_l), \end{aligned}$$

where $\theta^*_k = (\lambda^*_{1k}, \cdots, \lambda^*_{mk}, \beta^{*\prime}_k, \gamma^{*\prime}_k)'$ and

$$\begin{aligned} u^*_k(\theta^*_k) &= J(y_k - \textstyle\sum_{l=1}^m \lambda^*_{lk}\overline{y}_l - X\beta^*_k - \overline{X}\gamma^*_k) \\ &= J[\text{E}(\overline{y}_1), \cdots, \text{E}(\overline{y}_m), X, \overline{X}](\theta^*_{k,0} - \theta^*_k) + Ju^*_k + \textstyle\sum_{l=1}^m(\lambda^*_{lk,0} - \lambda^*_{lk})JG_l\text{vec}(U^*). \end{aligned}$$

As discussed above, the "pseudo" reduced form parameters can be asymptotically identified if $J[\text{E}(\overline{y}_1), \cdots, \text{E}(\overline{y}_m), X, \overline{X}]$ has full column rank for large enough n, where $\text{E}(\overline{y}_k) = G_k[(B_0^{*\prime} \otimes I_n + \Gamma_0^{*\prime} \otimes W)\text{vec}(X) + (I_m \otimes L)\text{vec}(\eta^*)]$. The term $G_k(I_m \otimes L)$ in $\text{E}(\overline{y}_k)$ 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 $JG_k(I_m \otimes L) = 0$. If $W_{(g)}$ has non-constant row sums for some g, then $G_k(I_m \otimes L)$ persists after the within transformation and provides additional information for identification (Liu, 2014).

When the rank condition fails, identification may still be possible through the quadratic moment conditions. The following proposition summarizes sufficient conditions for the "pseudo" reduced form parameters to be identified. Let σ_k^* denote the kth column of Σ^*.

Proposition 4.1. *$\theta_0^* = (\theta_{1,0}^{*\prime}, \cdots, \theta_{m,0}^{*\prime})'$ is asymptotically identified if either*

(i) *$J[\mathrm{E}(\overline{y}_1), \cdots, \mathrm{E}(\overline{y}_m), X, \overline{X}]$ has full column rank when n is sufficiently large; or*

(ii) *$J[\mathrm{E}(\overline{y}_1), \cdots, \mathrm{E}(\overline{y}_{\overline{m}}), X, \overline{X}]$ has full column rank for some $0 \leq \overline{m} \leq m-1$ when n is sufficiently large, and the equations*

$$\begin{aligned}&\lim_{n\to\infty} n^{-1}\{\textstyle\sum_{s=1}^m (\lambda_{sk,0}^* - \lambda_{sk}^*)\mathrm{tr}[\ddot{\Xi}_r' G_s(\sigma_l^* \otimes I_n)] + \sum_{t=1}^m (\lambda_{tl,0}^* - \lambda_{tl}^*)\mathrm{tr}[\ddot{\Xi}_r G_t(\sigma_k^* \otimes I_n)] \\ &+ \textstyle\sum_{s=1}^m \sum_{t=1}^m (\lambda_{sk,0}^* - \lambda_{sk}^*)(\lambda_{tl,0}^* - \lambda_{tl}^*)\mathrm{tr}[G_s' \ddot{\Xi}_r G_t(\Sigma^* \otimes I_n)]\} = 0, \end{aligned} \tag{4.8}$$

for $r = 1, \cdots, p$ and $k, l = 1, \cdots, m$, have a unique solution at Λ_0^*.

Identification of the structural parameters With the "pseudo" reduced form parameters $\Lambda_0^*, B_0^*, \Gamma_0^*$ identified, the structural parameters $\Theta_0 = [(I_m - \Phi_0)', -\Lambda_0', -B_0', -\Gamma_0']'$ can be identified through the linear restrictions (4.5) in the same way as in the classical linear simultaneous equations model. Suppose there are r_k exclusion restrictions of the form $R_k\vartheta_{k,0} = 0$ where R_k is a matrix of known constants and $\vartheta_{k,0}$ is the kth column of Θ_0. The sufficient and necessary *rank* condition for $\vartheta_{k,0}$ to be identified by the exclusion restrictions $R_k\vartheta_{k,0} = 0$ is that $\mathrm{rank}(R_k\Theta_0) = m-1$, and the necessary *order* condition is $r_k \geq m-1$. The following assumption summarizes the two-step identification strategy.

Assumption 9. The "pseudo" reduced-form parameters in (4.4) can be identified according to Proposition 4.1, and the structural parameters can be identified from the "pseudo" reduced-form parameters under proper exclusion restrictions.

Under the maintained assumptions, the following proposition establishes the consistency of the GMM estimator defined in (3.3).

Proposition 4.2. *Under Assumptions 1-9, if $q/n \to 0$ as $q, n \to \infty$, then $\widehat{\theta}_{gmm}$ is consistent.*

5 Asymptotic Normality

Let $V_k = JZ_k - F_k$ and $V = \text{diag}\{V_k\}$. Let $\Upsilon_{1,kl} = [\mathrm{E}(V_k'\ddot{\Xi}_1 u_l), \cdots, \mathrm{E}(V_k'\ddot{\Xi}_p u_l)]'$, $\Upsilon_{2,kl} = [\mathrm{E}(V_l'\ddot{\Xi}_1' u_k), \cdots, \mathrm{E}(V_l'\ddot{\Xi}_p' u_k)]'$,

$$D_2 = -\mathrm{E}[\frac{\partial}{\partial\theta'} h_2(\theta_0)] = \begin{bmatrix} \Upsilon_{1,11} & & \\ \vdots & & \\ \Upsilon_{1,1m} & & \\ & \ddots & \\ & & \Upsilon_{1,m1} \\ & & \vdots \\ & & \Upsilon_{1,mm} \end{bmatrix} + \begin{bmatrix} \Upsilon_{2,11} & & \\ & \ddots & \\ & & \Upsilon_{2,1m} \\ & \vdots & \\ \Upsilon_{2,m1} & & \\ & \ddots & \\ & & \Upsilon_{2,mm} \end{bmatrix},$$

and $D_2^* = D_2 - [(\mu_3'\Sigma^{-1}) \otimes \omega']F$. We maintain the following regularity condition.

Assumption 10. $\lim_{n\to\infty} n^{-1} D_2^*$ is a finite matrix with full column rank.

The following proposition gives the asymptotic distribution of the GMM estimator defined in (3.3).

Proposition 5.1. *Under Assumptions 1-10, if $q^{3/2}/n \to 0$ as $q, n \to \infty$, then*

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

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

The asymptotic covariance matrix of the GMM estimator can be compared with that of the 3SLS estimator in Liu (2014). The asymptotic covariance matrix of the 3SLS estimator is $\lim_{n\to\infty}[n^{-1}F'(\Sigma^{-1} \otimes I_n)F]^{-1}$. As $D_2^{*\prime}\Omega_{22}^{*-1}D_2^*$ is positive semi-definite, the GMM estimator improve the asymptotic efficiency of the 3SLS estimator.

The leading-order asymptotic bias of the GMM estimator given in Proposition 5.1 can be estimated to correct for the many-instrument bias. Suppose $\sqrt{n}\widehat{b}_{gmm}$ is a consistent estimator of $\sqrt{n}b_{gmm}$. The bias-corrected GMM (BCGMM) estimator is given by $\widehat{\theta}_{bcgmm} = \widehat{\theta}_{gmm} - \widehat{b}_{gmm}$. From Proposition 5.1, if $q^{3/2}/n \to 0$ then

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

In the following example, we derive the explicit form of the many-instrument bias for a simultaneous equations network model with $m = 2$.

Example 1. Suppose $m = 2$ and $k_x = 2$ with $X = [x_1, x_2]$. Consider the model

$$\begin{aligned} y_1 &= \phi_{21,0}y_2 + \lambda_{11,0}\overline{y}_1 + \lambda_{21,0}\overline{y}_2 + x_1\beta_{1,0} + \overline{x}_1\gamma_{1,0} + L\eta_1 + u_1 \\ y_2 &= \phi_{12,0}y_1 + \lambda_{12,0}\overline{y}_1 + \lambda_{22,0}\overline{y}_2 + x_2\beta_{2,0} + \overline{x}_2\gamma_{2,0} + L\eta_2 + u_2. \end{aligned} \tag{5.1}$$

Let

$$S = (1-\phi_{12,0}\phi_{21,0})I_n - (\lambda_{11,0}+\lambda_{22,0}+\phi_{12,0}\lambda_{21,0}+\phi_{21,0}\lambda_{12,0})W + (\lambda_{11,0}\lambda_{22,0}-\lambda_{12,0}\lambda_{21,0})W^2. \tag{5.2}$$

The reduced-form equations of model (5.1) are

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

where

$$\begin{aligned}
\mathrm{E}(y_1) &= S^{-1}[x_1\beta_{1,0} + Wx_1(\gamma_{1,0} - \lambda_{22,0}\beta_{1,0}) - W^2x_1\lambda_{22,0}\gamma_{1,0} \\
&\quad +x_2\phi_{21,0}\beta_{2,0} + Wx_2(\lambda_{21,0}\beta_{2,0} + \phi_{21,0}\gamma_{2,0}) + W^2x_2\lambda_{21,0}\gamma_{2,0} \\
&\quad +L(\eta_1 + \phi_{21,0}\eta_2) + WL(\lambda_{21,0}\eta_2 - \lambda_{22,0}\eta_1)] \\
\mathrm{E}(y_2) &= S^{-1}[x_2\beta_{2,0} + Wx_2(\gamma_{2,0} - \lambda_{11,0}\beta_{2,0}) - W^2x_2\lambda_{11,0}\gamma_{2,0} \\
&\quad +x_1\phi_{12,0}\beta_{1,0} + Wx_1(\lambda_{12,0}\beta_{1,0} + \phi_{12,0}\gamma_{1,0}) + W^2x_1\lambda_{12,0}\gamma_{1,0} \\
&\quad +L(\eta_2 + \lambda_{12,0}\eta_1) + WL(\lambda_{12,0}\eta_1 - \lambda_{11,0}\eta_2)]
\end{aligned}$$

and

$$\begin{aligned}
\epsilon_1 &= (I_n - \lambda_{22,0}W)S^{-1}u_1 + (\phi_{21,0}I_n + \lambda_{21,0}W)S^{-1}u_2 \\
\epsilon_2 &= (I_n - \lambda_{11,0}W)S^{-1}u_2 + (\phi_{12,0}I_n + \lambda_{12,0}W)S^{-1}u_1.
\end{aligned}$$

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

$$\begin{aligned}
F_1 &= \mathrm{E}(JZ_1) = J[\mathrm{E}(y_2), W\mathrm{E}(y_1), W\mathrm{E}(y_2), x_1, \overline{x}_1] \\
F_2 &= \mathrm{E}(JZ_2) = J[\mathrm{E}(y_1), W\mathrm{E}(y_1), W\mathrm{E}(y_2), x_2, \overline{x}_2]
\end{aligned}$$

and

$$\begin{aligned}
V_1 &= JZ_1 - F_1 = J[\epsilon_2, W\epsilon_1, W\epsilon_2, 0_{n\times 2}] \qquad (5.3) \\
V_2 &= JZ_2 - F_2 = J[\epsilon_1, W\epsilon_1, W\epsilon_2, 0_{n\times 2}].
\end{aligned}$$

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

$$h_1(\theta) = (I_2 \otimes \ddot{Q})'[u_1(\theta_1)', u_2(\theta_2)']',$$

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

$$h_2(\theta) = [h_{2,11}(\theta_1, \theta_1)', h_{2,12}(\theta_1, \theta_2)', h_{2,21}(\theta_2, \theta_1)', h_{2,22}(\theta_2, \theta_2)']',$$

where $h_{2,kl}(\theta_k, \theta_l) = [\ddot{\Xi}_1' u_k(\theta_k), \cdots, \ddot{\Xi}_p' u_k(\theta_k)]' u_l(\theta_l)$ for $k, l = 1, 2$. Let $\mu_{s,t} = \mathrm{E}(u_{i1}^s u_{i2}^t)$, for $s + t = 3, 4$. Then,

$$\mu_3 = \begin{bmatrix} \mu_{3,0} & \mu_{2,1} & \mu_{2,1} & \mu_{1,2} \\ \mu_{2,1} & \mu_{1,2} & \mu_{1,2} & \mu_{0,3} \end{bmatrix},$$

and

$$\begin{aligned} \Omega_{22} &= \kappa_4 \otimes (\omega'\omega) \\ &+ \begin{bmatrix} \sigma_{11}^2 & \sigma_{11}\sigma_{12} & \sigma_{11}\sigma_{12} & \sigma_{12}^2 \\ * & \sigma_{12}^2 & \sigma_{11}\sigma_{22} & \sigma_{12}\sigma_{22} \\ * & * & \sigma_{12}^2 & \sigma_{12}\sigma_{22} \\ * & * & * & \sigma_{22}^2 \end{bmatrix} \otimes \Delta_1 + \begin{bmatrix} \sigma_{11}^2 & \sigma_{11}\sigma_{12} & \sigma_{11}\sigma_{12} & \sigma_{12}^2 \\ * & \sigma_{11}\sigma_{22} & \sigma_{12}^2 & \sigma_{12}\sigma_{22} \\ * & * & \sigma_{11}\sigma_{22} & \sigma_{12}\sigma_{22} \\ * & * & * & \sigma_{22}^2 \end{bmatrix} \otimes \Delta_2, \end{aligned}$$

where $\omega = [\mathrm{vec}_D(\ddot{\Xi}_1), \cdots, \mathrm{vec}_D(\ddot{\Xi}_p)]$,

$$\kappa_4 = \begin{bmatrix} \mu_{4,0} - 3\sigma_{11}^2 & \mu_{3,1} - 3\sigma_{11}\sigma_{12} & \mu_{3,1} - 3\sigma_{11}\sigma_{12} & \mu_{2,2} - \sigma_{11}\sigma_{22} - 2\sigma_{12}^2 \\ * & \mu_{2,2} - \sigma_{11}\sigma_{12} - 2\sigma_{12}^2 & \mu_{2,2} - \sigma_{11}\sigma_{12} - 2\sigma_{12}^2 & \mu_{1,3} - 3\sigma_{12}\sigma_{22} \\ * & * & \mu_{2,2} - \sigma_{11}\sigma_{22} - 2\sigma_{12}^2 & \mu_{1,3} - 3\sigma_{12}\sigma_{22} \\ * & * & * & \mu_{0,4} - 3\sigma_{22}^2 \end{bmatrix},$$

$$\Delta_1 = \begin{bmatrix} \mathrm{tr}(\ddot{\Xi}_1\ddot{\Xi}_1) & \cdots & \mathrm{tr}(\ddot{\Xi}_1\ddot{\Xi}_p) \\ \vdots & \ddots & \vdots \\ \mathrm{tr}(\ddot{\Xi}_p\ddot{\Xi}_1) & \cdots & \mathrm{tr}(\ddot{\Xi}_p\ddot{\Xi}_p) \end{bmatrix}, \quad \text{and} \quad \Delta_2 = \begin{bmatrix} \mathrm{tr}(\ddot{\Xi}_1\ddot{\Xi}_1') & \cdots & \mathrm{tr}(\ddot{\Xi}_1\ddot{\Xi}_p') \\ \vdots & \ddots & \vdots \\ \mathrm{tr}(\ddot{\Xi}_p\ddot{\Xi}_1') & \cdots & \mathrm{tr}(\ddot{\Xi}_p\ddot{\Xi}_p') \end{bmatrix}.$$

Let $\widehat{h}_2^*(\theta) = h_2(\theta) - [(\widehat{\mu}_3'\widehat{\Sigma}^{-1}) \otimes (\omega'\ddot{P})]u(\theta)$ and $\widehat{\Omega}_{22}^* = \widehat{\Omega}_{22} - (\widehat{\mu}_3'\widehat{\Sigma}^{-1}\widehat{\mu}_3) \otimes (\omega'\ddot{P}\omega)$, where $\widehat{\Sigma}$, $\widehat{\mu}_3$ and $\widehat{\Omega}_{22}$ are $\sqrt{n}$-consistent estimators of Σ, μ_3 and Ω_{22}.[7] Then, it follows from Proposition 5.1 that, if $q^{3/2}/n \to 0$, the GMM estimator $\widehat{\theta}_{gmm} = \arg\min u(\theta)'(\widehat{\Sigma}^{-1} \otimes \ddot{P})u(\theta) + \widehat{h}_2^*(\theta)'\widehat{\Omega}_{22}^{*-1}\widehat{h}_2^*(\theta)$ is asymptotically normal with an asymptotic bias

$$b_{gmm} = [F'(\Sigma^{-1} \otimes I_n)F + D_2^{*\prime}\Omega_{22}^{*-1}D_2^*]^{-1}\mathrm{E}[V'(\Sigma^{-1} \otimes \ddot{P})u], \tag{5.4}$$

[7] When $u = (u_1', u_2')' \sim N(0, \Sigma \otimes I_n)$, $\widehat{h}_2^*(\theta)$ and $\widehat{\Omega}_{22}^*$ can be simplified as $\mu_3 = 0$ and $\kappa_4 = 0$.

where

$$\mathrm{E}[V'(\Sigma^{-1}\otimes\ddot{P})u]=\begin{bmatrix}\phi_{12,0}\mathrm{tr}(\ddot{P}S^{-1})+\lambda_{12,0}\mathrm{tr}(\ddot{P}WS^{-1})\\ \mathrm{tr}(\ddot{P}WS^{-1})-\lambda_{22,0}\mathrm{tr}(\ddot{P}W^2S^{-1})\\ \phi_{12,0}\mathrm{tr}(\ddot{P}WS^{-1})+\lambda_{12,0}\mathrm{tr}(\ddot{P}W^2S^{-1})\\ 0_{2\times1}\\ \phi_{21,0}\mathrm{tr}(\ddot{P}S^{-1})+\lambda_{21,0}\mathrm{tr}(\ddot{P}WS^{-1})\\ \phi_{21,0}\mathrm{tr}(\ddot{P}WS^{-1})+\lambda_{21,0}\mathrm{tr}(\ddot{P}W^2S^{-1})\\ \mathrm{tr}(\ddot{P}WS^{-1})-\lambda_{11,0}\mathrm{tr}(\ddot{P}W^2S^{-1})\\ 0_{2\times1}\end{bmatrix}. \tag{5.5}$$

Let $\widehat{\mathrm{E}}[V'(\Sigma^{-1}\otimes\ddot{P})u]$ denote the estimated $\mathrm{E}[V'(\Sigma^{-1}\otimes\ddot{P})u]$ with $\phi_{lk,0}$'s and $\lambda_{lk,0}$'s in (5.5) replaced by their $\sqrt{n}$-consistent preliminary estimates.[8] Let $\widehat{D}_2^*=\widehat{D}_2-[(\widehat{\mu}_3'\widehat{\Sigma}^{-1})\otimes\omega'\ddot{P}]Z$, where $n^{-1}\widehat{D}_2$ is a consistent estimator of $n^{-1}D_2$.[9] If follows by Lemma A.1 in the appendix that, if $q^{3/2}/n\to0$, the BCGMM estimator $\widehat{\theta}_{bcgmm}=\widehat{\theta}_{gmm}-\widehat{b}_{gmm}$, where

$$\widehat{b}_{gmm}=[Z'(\Sigma^{-1}\otimes\ddot{P})Z+\widehat{D}_2^{*\prime}\widehat{\Omega}_{22}^{*-1}\widehat{D}_2^*]^{-1}\widehat{\mathrm{E}}[V'(\Sigma^{-1}\otimes\ddot{P})u], \tag{5.6}$$

has an asymptotically normal distribution around θ_0. □

6 Monte Carlo Experiments

To investigate the finite sample performance of the proposed GMM estimator, we conduct a limited simulation study based on model (5.1). 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 ith row of $W_{(g)}$, we generate an integer $c_{g,i}$ uniformly at random from the set

[8] For example, Φ_0 and Λ_0 can be consistently estimated by a less efficient equation-by-equation 2SLS estimator with a fixed number of IVs $\ddot{Q}=J[X,WX,W^2X]$.

[9] The explicit expression of $n^{-1}D_2$ and its estimator can be found in the proof of Lemma A.1.

of integers $\{1,2,3\}$. Then, if $i+c_{g,i} \leq n_g$, we set the $(i+1)$th, $\cdots$, $(i+c_{g,i})$th elements of the ith 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 ith row will be ones. We experiment with different numbers of groups $\bar{g}$ 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 η_k is generated from $N(0, I_{\bar{g}})$ for $k = 1, 2$. The error term $u = (u_1', u_2')'$ is generated from $N(0, \Sigma \otimes I_n)$. We set $\sigma_{11} = \sigma_{22} = 1$, $\phi_{21,0} = \phi_{12,0} = 0.2$, and $\lambda_{11,0} = \lambda_{21,0} = \lambda_{12,0} = \lambda_{22,0} = 0.1$.[10] We experiment with different values for σ_{12}, $(\beta_{1,0}, \beta_{2,0})$ 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 $\ddot{Q}_1 = J[X, WX, W^2X]$, where $X = [x_1, x_2]$; (ii) 3SLS-2: the 3SLS estimator with the IV matrix $\ddot{Q}_2 = [\ddot{Q}_1, JWL]$; (iii) BC3SLS: the bias-corrected 3SLS-2; (iv) GMM-1: the GMM estimator with the IV matrix $\ddot{Q}_1$ and quadratic moment functions $h_2(\theta) = [u_1(\theta_1)'\ddot{\Xi}'u_1(\theta_1), u_2(\theta_2)'\ddot{\Xi}'u_1(\theta_1), u_1(\theta_1)'\ddot{\Xi}'u_2(\theta_2), u_2(\theta_2)'\ddot{\Xi}'u_2(\theta_2)]'$, where $\ddot{\Xi} = JWJ - \mathrm{tr}(JW)J/\mathrm{tr}(J)$; (v) GMM-2: the GMM estimator with the IV matrix $\ddot{Q}_2$ and the same set of quadratic moment functions used by GMM-1; and (vi) BCGMM: the bias-corrected GMM-2. The IV matrix $\ddot{Q}_1$ is based on the exogenous attributes of direct and indirect connections. $\ddot{Q}_2$ includes additional IVs JWL based on the numbers of (direct) connections to improve estimation efficiency. As WL has $\bar{g}$ columns, the number of IVs in $\ddot{Q}_2$ 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

[10] The matrix S defined in (5.2) is invertible with the chosen parameters.

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 (5.1), we only report the estimation result for the first equation of model (5.1) 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 $\bar{g} = 30$ 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 $\ddot{Q}_1$ 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 $\bar{g} = 30$ 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 $\ddot{Q}_2$ 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 $\bar{g} = 30$ 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 $\ddot{Q}_1$ 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 $\bar{g} = 30$ 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 $\ddot{Q}_2$ introduce biases into the estimators. The size of the bias increases as the correlation across equations σ_{12} increases and as $\ddot{Q}_1$ 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.

7 Summary and Future Work

In this paper, we propose a new set of quadratic moment conditions based on the correlation of individual decisions in multiple activities to identify peer effects. Combining linear and quadratic moment conditions, we develop a general GMM framework to estimate the simultaneous equations network model. 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 study activity-specific networks and associated identification problems.

References

Bekker, P. (1994). Alternative approximations to the distributions of instrumental variable estimators, *Econometrica* **62**: 657–681.

Blume, L. E., Brock, W. A., Durlauf, S. N. and Ioannides, Y. M. (2011). Identification of social interactions, *in* J. Benhabib, A. Bisin and M. O. Jackson (eds), *Handbook of Social Economics*, Vol. 1B, North-Holland, pp. 855–966.

Bonacich, P. (1987). Power and centrality: a family of measures, *American Journal of Sociology* **92**: 1170–1182.

Bramoullé, Y., Djebbari, H. and Fortin, B. (2009). Identification of peer effects through social networks, *Journal of Econometrics* **150**: 41–55.

Cohen-Cole, E., Liu, X. and Zenou, Y. (2018). Multivariate choices and identification of social interactions, *Journal of Applied Econometrics* **33**: 165–178.

Donald, S. G. and Newey, W. K. (2001). Choosing the number of instruments, *Econometrica* **69**: 1161–1191.

Durlauf, S. N. and Tanaka, H. (2008). Understanding regression versus variance tests for social interactions, *Economic Inquiry* **46**: 25–28.

Glaeser, E., Sacerdote, B. and Scheinkman, J. (1996). Crime and social interactions, *The Quarterly Journal of Economics* **111**: 507–548.

Graham, B. S. (2008). Identifying social interactions through conditional variance restrictions, *Econometrica* **76**: 643–660.

Hansen, C., Hausman, J. and Newey, W. (2008). Estimation with many instrumental variables, *Journal of Business and Economic Statistics* **26**: 398–422.

Heckman, J. J. (1976). The common structure of statistical models of truncation, sample selection and limited dependent variables and a simple estimator of such models, *Annals of Economic and Social Measurement* **5**: 475–492.

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

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

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

Liu, X. (2014). Identification and efficient estimation of simultaneous equations network models, *Journal of Business & Economic Statistics* **32**: 516–536.

Liu, X. and Lee, L. F. (2010). GMM estimation of social interaction models with centrality, *Journal of Econometrics* **159**: 99–115.

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

Liu, X. and Saraiva, P. (2015). Gmm estimation of sar models with endogenous regressors, *Regional Science and Urban Economics* **55**: 68–79.

Liu, X. and Saraiva, P. (2019). Gmm estimation of spatial autoregressive models in a system of simultaneous equations with heteroskedasticity, *Econometric Reviews* **38**: 359–385.

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

Neyman, J. and Scott, E. L. (1948). Consistent estimates based on partially consistent observations, *Econometrica* **16**: 1–32.

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

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

A Proofs

Proof of Proposition 4.1. If condition (i) holds, then θ_0^* is asymptotically identified from (4.1). Condition (i) fails if, for some $0 \leq \overline{m} \leq m-1$, $J\mathrm{E}(\overline{y}_l)$ is linearly dependent on $J[\mathrm{E}(\overline{y}_1), \cdots, \mathrm{E}(\overline{y}_{\overline{m}}), X, \overline{X}]$ such that $J\mathrm{E}(\overline{y}_l) = b_{l,1}J\mathrm{E}(\overline{y}_1) + \cdots + b_{l,\overline{m}}J\mathrm{E}(\overline{y}_{\overline{m}}) + JXc_{l,1} + J\overline{X}c_{l,2}$, for $l = \overline{m}+1, \cdots, m$, where $b_{l,1}, \cdots, b_{l,\overline{m}}$ are constant scalars and $c_{l,1}, c_{l,2}$ are constant vectors. If $J[\mathrm{E}(\overline{y}_1), \cdots, \mathrm{E}(\overline{y}_{\overline{m}}), X, \overline{X}]$ has full column rank for large enough n, (4.1) implies

$$\begin{aligned}
\lambda_{1k}^* &= \lambda_{1k,0}^* + \textstyle\sum_{l=\overline{m}+1}^{m}(\lambda_{lk,0}^* - \lambda_{lk}^*)b_{l,1}, \qquad \text{(A.1)} \\
& \cdots \\
\lambda_{\overline{m}k}^* &= \lambda_{\overline{m}k,0}^* + \textstyle\sum_{l=\overline{m}+1}^{m}(\lambda_{lk,0}^* - \lambda_{lk}^*)b_{l,\overline{m}}, \\
\beta_k^* &= \beta_{k,0}^* + \textstyle\sum_{l=\overline{m}+1}^{m}(\lambda_{lk,0}^* - \lambda_{lk}^*)c_{l,1}, \\
\gamma_k^* &= \gamma_{k,0}^* + \textstyle\sum_{l=\overline{m}+1}^{m}(\lambda_{lk,0}^* - \lambda_{lk}^*)c_{l,2},
\end{aligned}$$

for $k = 1, \cdots, m$, i.e., θ_0^* is identified if Λ_0^* is identified. Substitution of (A.1) into (4.3) gives (4.8). Therefore, Λ_0^* is identified if (4.8) has a unique solution at Λ_0^*. The desired result follows. □

Proof of Proposition 4.2. First, we consider the infeasible GMM estimator

$$\widetilde{\theta}_{gmm} = \arg\min h(\theta)'\Omega^{-1}h(\theta) = \arg\min u(\theta)'(\Sigma^{-1} \otimes \ddot{P})u(\theta) + h_2^*(\theta)'\Omega_{22}^{*-1}h_2^*(\theta).$$

Let $V_k = JZ_k - F_k$ and $V = \mathrm{diag}\{V_k\}$. $u(\theta) = J(y - Z\theta) = d(\theta) + r(\theta)$, where $d(\theta) = F(\theta_0 - \theta)$ and $r(\theta) = V(\theta_0 - \theta) + Ju$. Suppose $F(\theta_0 - \theta) = F^*(\theta_0^{(1)} - \theta^{(1)}) + F^*C(\theta_0^{(2)} - \theta^{(2)})$ for some constant matrix C, where $\theta_0^{(1)}$ is a vector of coefficients corresponding to the linearly independent columns of F. When $q/n \to 0$, it follows by Lemma C.3 of Liu (2014) that

$$\begin{aligned}
& n^{-1}u(\theta)'(\Sigma^{-1} \otimes \ddot{P})u(\theta) \\
= \; & n^{-1}d(\theta)'(\Sigma^{-1} \otimes \ddot{P})d(\theta) + 2n^{-1}d(\theta)'(\Sigma^{-1} \otimes \ddot{P})r(\theta) + n^{-1}r(\theta)'(\Sigma^{-1} \otimes \ddot{P})r(\theta) \\
= \; & n^{-1}(\theta_0 - \theta)'F'(\Sigma^{-1} \otimes I_n)F(\theta_0 - \theta) + o_p(1) \\
= \; & n^{-1}[(\theta_0^{(1)} - \theta^{(1)}) + C(\theta_0^{(2)} - \theta^{(2)})]'F^{*\prime}(\Sigma^{-1} \otimes I_n)F^*[(\theta_0^{(1)} - \theta^{(1)}) + C(\theta_0^{(2)} - \theta^{(2)})] + o_p(1),
\end{aligned}$$

and

$$\begin{aligned}
& n^{-1}[(\mu_3'\Sigma^{-1}) \otimes (\omega'\ddot{P})]u(\theta) \\
= \; & n^{-1}[(\mu_3'\Sigma^{-1}) \otimes (\omega'\ddot{P})]d(\theta) + n^{-1}[(\mu_3'\Sigma^{-1}) \otimes (\omega'\ddot{P})]r(\theta) \\
= \; & n^{-1}[(\mu_3'\Sigma^{-1}) \otimes \omega']F(\theta_0 - \theta) + o_p(1) \\
= \; & n^{-1}[(\mu_3'\Sigma^{-1}) \otimes \omega']F^*[(\theta_0^{(1)} - \theta^{(1)}) + C(\theta_0^{(2)} - \theta^{(2)})] + o_p(1),
\end{aligned}$$

uniformly in θ. In addition, it follows by a similar argument as in the proof of Proposition 1 in Lee (2007) that $n^{-1}h_2(\theta) - n^{-1}\mathrm{E}[h_2(\theta)] = o_p(1)$ uniformly in θ. Hence, $n^{-1}h_2^*(\theta) - n^{-1}\overline{h}_2^*(\theta) = o_p(1)$ uniformly in θ. As $n^{-1}\overline{h}_2^*(\theta)$ is a quadratic function of θ and the parameter space of θ is bounded, $n^{-1}\overline{h}_2^*(\theta)$ is uniformly equicontinuous in θ. The identification condition and uniform equicontinuity of $n^{-1}\overline{h}_2^*(\theta)$ imply that the identification uniqueness condition for $n^{-1}\overline{h}_2^*(\theta)\Omega_{22}^{*-1}\overline{h}_2^*(\theta)$ must be satisfied. The consistency of $\widetilde{\theta}_{gmm}$ follows from the uniform convergence and identification uniqueness condition for $\mathrm{plim}_{n\to\infty} n^{-1}[u(\theta)'(\Sigma^{-1} \otimes \ddot{P})u(\theta) + h_2^*(\theta)'\Omega_{22}^{*-1}h_2^*(\theta)]$ (White, 1994).

Let $\widehat{\Sigma}$ and $\widehat{\mu}_3$ be $\sqrt{n}$-consistent estimators of Σ and μ_3 respectively. Let $\widehat{h}_2^*(\theta) = h_2(\theta) - [(\widehat{\mu}_3'\widehat{\Sigma}^{-1}) \otimes (\omega'\ddot{P})]u(\theta)$ and $\widehat{\Omega}_{22}^* = \widehat{\Omega}_{22} - (\widehat{\mu}_3'\widehat{\Sigma}^{-1}\widehat{\mu}_3) \otimes (\omega'\ddot{P}\omega)$, where $n^{-1}\widehat{\Omega}_{22}$ is a $\sqrt{n}$-consistent estimator of $n^{-1}\Omega_{22}$. It remains to show that $n^{-1}u(\theta)'[(\widehat{\Sigma}^{-1} - \Sigma^{-1}) \otimes \ddot{P}]u(\theta) =$

$o_p(1)$ and $n^{-1}\widehat{h}_2^*(\theta)'\widehat{\Omega}_{22}^{*-1}\widehat{h}_2^*(\theta) - n^{-1}h_2^*(\theta)'\Omega_{22}^{*-1}h_2^*(\theta) = o_p(1)$ uniformly in θ. By a similar argument as above, when $q/n \to 0$,

$$n^{-1}u(\theta)'[(\widehat{\Sigma}^{-1} - \Sigma^{-1}) \otimes \ddot{P}]u(\theta) = n^{-1}(\theta_0 - \theta)'F'[(\widehat{\Sigma}^{-1} - \Sigma^{-1}) \otimes I_n]F(\theta_0 - \theta) + o_p(1),$$

uniformly in θ. As $\widehat{\Sigma}^{-1} - \Sigma^{-1} = o_p(1)$, $n^{-1}u(\theta)'[(\widehat{\Sigma}^{-1} - \Sigma^{-1}) \otimes \ddot{P}]u(\theta) = o_p(1)$ uniformly in θ. On the other hand,

$$\begin{aligned}
& n^{-1}\widehat{h}_2^*(\theta)'\widehat{\Omega}_{22}^{*-1}\widehat{h}_2^*(\theta) - n^{-1}h_2^*(\theta)'\Omega_{22}^{*-1}h_2^*(\theta) \\
= \; & n^{-1}[\widehat{h}_2^*(\theta) - h_2^*(\theta)]'(n^{-1}\widehat{\Omega}_{22}^*)^{-1}n^{-1}\widehat{h}_2^*(\theta) + n^{-1}h_2^*(\theta)'(n^{-1}\widehat{\Omega}_{22}^*)^{-1}n^{-1}[\widehat{h}_2^*(\theta) - h_2^*(\theta)] \\
& + n^{-1}h_2^*(\theta)'[(n^{-1}\widehat{\Omega}_{22}^*)^{-1} - (n^{-1}\Omega_{22}^*)^{-1}]n^{-1}h_2^*(\theta).
\end{aligned}$$

As shown above, $n^{-1}h_2^*(\theta) - n^{-1}\overline{h}_2^*(\theta) = o_p(1)$ uniformly in θ. By a similar argument as in the proof of Proposition 2 in Lee (2007), $n^{-1}\mathrm{E}[h_2(\theta)] = O(1)$ uniformly in θ. Therefore, $n^{-1}\mathrm{E}[h_2^*(\theta)] = O(1)$ uniformly in θ, which implies $n^{-1}h_2^*(\theta) = O_p(1)$ uniformly in θ.

$$\begin{aligned}
& n^{-1}[\widehat{h}_2^*(\theta) - h_2^*(\theta)] \\
= \; & -n^{-1}[(\widehat{\mu}_3'\widehat{\Sigma}^{-1} - \mu_3'\Sigma^{-1}) \otimes (\omega'\ddot{P})]u(\theta) \\
= \; & -n^{-1}[(\widehat{\mu}_3'\widehat{\Sigma}^{-1} - \mu_3'\Sigma^{-1}) \otimes (\omega'\ddot{P})]F(\theta_0 - \theta) - n^{-1}[(\widehat{\mu}_3'\widehat{\Sigma}^{-1} - \mu_3'\Sigma^{-1}) \otimes (\omega'\ddot{P})]r(\theta).
\end{aligned}$$

When $q/n \to 0$, it follows by Lemma C.3 of Liu (2014) that $n^{-1}[\widehat{h}_2^*(\theta) - h_2^*(\theta)] = o_p(1)$ uniformly in θ. As $n^{-1}\widehat{\Omega}_{22}^* - n^{-1}\Omega_{22}^* = o_p(1)$, we have $n^{-1}\widehat{h}_2^*(\theta)'\widehat{\Omega}_{22}^{*-1}\widehat{h}_2^*(\theta) - n^{-1}h_2^*(\theta)'\Omega_{22}^{*-1}h_2^*(\theta) = o_p(1)$ uniformly in θ. The desired result follows. □

Proof of Proposition 5.1. The Taylor expansion of

$$-Z'(\widehat{\Sigma}^{-1} \otimes \ddot{P})u(\widehat{\theta}_{gmm}) + \frac{\partial \widehat{h}_2^*(\widehat{\theta}_{gmm})'}{\partial \theta}\widehat{\Omega}_{22}^{*-1}\widehat{h}_2^*(\widehat{\theta}_{gmm}) = 0$$

around θ_0 gives $\sqrt{n}(\widehat{\theta}_{gmm} - \theta_0) = \widehat{A}^{-1}\widehat{b}$, where

$$\begin{aligned}
\widehat{A} &= n^{-1}Z'(\widehat{\Sigma}^{-1} \otimes \ddot{P})Z + n^{-1}\frac{\partial \widehat{h}_2^*(\widehat{\theta}_{gmm})'}{\partial \theta}\widehat{\Omega}_{22}^{*-1}\frac{\partial \widehat{h}_2^*(\theta^+)}{\partial \theta'}, \\
\widehat{b} &= n^{-1/2}Z'(\widehat{\Sigma}^{-1} \otimes \ddot{P})u - n^{-1/2}\frac{\partial \widehat{h}_2^*(\widehat{\theta}_{gmm})'}{\partial \theta}\widehat{\Omega}_{22}^{*-1}\widehat{h}_2^*(\theta_0),
\end{aligned}$$

for some θ^+ between $\widehat{\theta}_{gmm}$ and θ_0. As $\widehat{\Sigma} - \Sigma = O_p(n^{-1/2})$, it follows by Lemma C.3 of Liu (2014) that, if $q/n \to 0$,

$$\begin{aligned}
n^{-1}Z'(\widehat{\Sigma}^{-1} \otimes \ddot{P})Z &= n^{-1}F'(\widehat{\Sigma}^{-1} \otimes I_n)F - n^{-1}F'[\widehat{\Sigma}^{-1} \otimes (I_n - \ddot{P})]F \\
&\quad + n^{-1}F'(\widehat{\Sigma}^{-1} \otimes \ddot{P})V + n^{-1}V'(\widehat{\Sigma}^{-1} \otimes \ddot{P})F + n^{-1}V'(\widehat{\Sigma}^{-1} \otimes \ddot{P})V \\
&= n^{-1}F'(\Sigma^{-1} \otimes I_n)F + O_p(\sqrt{q/n}).
\end{aligned}$$

$\partial \widehat{h}_2^*(\theta)/\partial \theta' = \partial h_2(\theta)/\partial \theta' + [(\widehat{\mu}_3'\widehat{\Sigma}^{-1}) \otimes (\omega'\ddot{P})]Z$. For a typical element of $h_2(\theta)$, we have

$$\begin{aligned}
&\frac{\partial u_k(\theta_k)'}{\partial \theta_k}\ddot{\Xi}_r u_l(\theta_l) = -Z_k'\ddot{\Xi}_r u_l(\theta_l) \qquad \text{(A.2)} \\
= \; & -F_k'\ddot{\Xi}_r u_l - V_k'\ddot{\Xi}_r u_l + (F_k'\ddot{\Xi}_r F_l + F_k'\ddot{\Xi}_r V_l + V_k'\ddot{\Xi}_r F_l + V_k'\ddot{\Xi}_r V_l)(\theta_l - \theta_{l,0}).
\end{aligned}$$

It follows from Lemmas A.4 and A.5 of Lee (2007) that $n^{-1}\partial h_2(\widetilde{\theta})/\partial \theta' = -n^{-1}D_2 + o_p(1)$ for $\widetilde{\theta} = \theta_0 + o_p(1)$. As $\widehat{\Sigma} - \Sigma = O_p(n^{-1/2})$ and $\widehat{\mu}_3 - \mu_3 = O_p(n^{-1/2})$, it follows by Lemma C.3 of Liu (2014) that, if $q/n \to 0$,

$$\begin{aligned}
&n^{-1}[(\widehat{\mu}_3'\widehat{\Sigma}^{-1}) \otimes (\omega'\ddot{P})]Z \\
= \; & n^{-1}[(\widehat{\mu}_3'\widehat{\Sigma}^{-1}) \otimes \omega']F - n^{-1}\{(\widehat{\mu}_3'\widehat{\Sigma}^{-1}) \otimes [\omega'(I_n - \ddot{P})]\}F + n^{-1}[(\widehat{\mu}_3'\widehat{\Sigma}^{-1}) \otimes (\omega'\ddot{P})]V \\
= \; & n^{-1}[(\mu_3'\Sigma^{-1}) \otimes \omega']F + O_p(\sqrt{q/n}).
\end{aligned}$$

Therefore, $n^{-1}\partial \widehat{h}_2^*(\widetilde{\theta})/\partial \theta' = -n^{-1}D_2^* + o_p(1) = O_p(1)$ for $\widetilde{\theta} = \theta_0 + o_p(1)$, which implies that

$n^{-1}\frac{\partial}{\partial\theta}\widehat{h}_2^*(\widehat{\theta}_{gmm})'\widehat{\Omega}_{22}^{*-1}\frac{\partial}{\partial\theta'}\widehat{h}_2^*(\theta^+) - n^{-1}D_2^{*\prime}\Omega_{22}^{*-1}D_2^* = o_p(1)$ since $n^{-1/2}(\widehat{\Omega}_{22}^* - \Omega_{22}^*) = O_p(1)$. In summary,

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

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

$$\begin{aligned}
& n^{-1/2}Z'(\widehat{\Sigma}^{-1}\otimes\ddot{P})u \\
= \; & n^{-1/2}F'(\widehat{\Sigma}^{-1}\otimes I_n)u - n^{-1/2}F'[\widehat{\Sigma}^{-1}\otimes(I_n - \ddot{P})]u + n^{-1/2}V'(\widehat{\Sigma}^{-1}\otimes\ddot{P})u \\
= \; & n^{-1/2}F'(\Sigma^{-1}\otimes I_n)u + n^{-1/2}\mathrm{E}[V'(\Sigma^{-1}\otimes\ddot{P})u] + o_p(1),
\end{aligned}$$

and $n^{-1/2}\widehat{h}_2^*(\theta_0) - n^{-1/2}h_2^*(\theta_0) = -n^{-1}[\sqrt{n}(\widehat{\mu}_3'\widehat{\Sigma}^{-1} - \mu_3'\Sigma^{-1})\otimes(\omega'\ddot{P})]u = o_p(1)$. As

$$n^{-1}\mathrm{E}[h_2^*(\theta_0)u'(\Sigma^{-1}\otimes I_n)F] = n^{-1}\{(\mu_3'\Sigma^{-1})\otimes[\omega'(I_n - \ddot{P})]\}F = o_p(1),$$

$n^{-1/2}F'(\Sigma^{-1}\otimes I_n)u$ and $n^{-1/2}h_2^*(\theta_0)$ are asymptotically uncorrelated. It follows by Lemma 3 of Yang and Lee (2017) that

$$\widehat{b} - n^{-1/2}\mathrm{E}[V'(\Sigma^{-1}\otimes\ddot{P})u] \xrightarrow{d} N(0, \lim_{n\to\infty} n^{-1}[F'(\Sigma^{-1}\otimes I_n)F + D_2^{*\prime}\Omega_{22}^{*-1}D_2^*]). \tag{A.4}$$

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

It follows from (A.2) and Lemmas A.4 and A.5 of Lee (2007) that, if $q/n \to 0$, we have $n^{-1}\partial h_2(\widetilde{\theta})/\partial\theta' = -n^{-1}D_2 + O_p(\max\{1/\sqrt{n}, q/n\})$ for $\widetilde{\theta} - \theta_0 = O_p(q/n)$, which implies that $n^{-1}\partial h_2^*(\widetilde{\theta})/\partial\theta' = -n^{-1}D_2^* + O_p(\sqrt{q/n})$. Hence,

$$\widehat{A} = n^{-1}[F'(\Sigma^{-1}\otimes I_n)F + D_2^{*\prime}\Omega_{22}^{*-1}D_2^*] + O_p(\sqrt{q/n}). \tag{A.5}$$

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

$$\begin{aligned}
&\sqrt{n}(\widehat{\theta}_{gmm} - \theta_0 - b_{gmm}) \\
= \quad & -[n^{-1}\frac{\partial h(\widehat{\theta}_{gmm})'}{\partial \theta}\Omega^{-1}\frac{\partial h(\theta^{+})}{\partial \theta'}]^{-1}n^{-1/2}\{\frac{\partial h(\widehat{\theta}_{gmm})'}{\partial \theta}\Omega^{-1}h(\theta_0) + \mathrm{E}[V'(\Sigma^{-1}\otimes \ddot{P})u]\} + O_p(\sqrt{q^3/n^2}).
\end{aligned}$$

Hence, if $q^{3/2}/n \to 0$, $\sqrt{n}(\widehat{\theta}_{gmm} - \theta_0 - b_{gmm}) \xrightarrow{d} N(0, \lim_{n\to\infty} n[F'(\Sigma^{-1}\otimes I_n)F + D_2^{*\prime}\Omega_{22}^{*-1}D_2^*]^{-1})$.

□

Lemma A.1. *If $q/n \to 0$ then $\sqrt{n}(\widehat{b}_{gmm} - b_{gmm}) = o_p(1)$, where b_{gmm} and $\widehat{b}_{gmm}$ are given by (5.4) and (5.6) respectively.*

Proof. To show the desired result, it is sufficient to show that $n^{-1}Z'(\widehat{\Sigma}^{-1}\otimes \ddot{P})Z - n^{-1}F'(\Sigma^{-1}\otimes I_n)F = o_p(1)$, $n^{-1}\widehat{D}_2^* - n^{-1}D_2^* = o_p(1)$, $n^{-1}\widehat{\Omega}_{22}^* - n^{-1}\Omega_{22}^* = o_p(1)$, and $n^{-1/2}\widehat{\mathrm{E}}[V'(\Sigma^{-1}\otimes \ddot{P})u] - n^{-1/2}\mathrm{E}[V'(\Sigma^{-1}\otimes \ddot{P})u] = o_p(1)$. By a similar argument as in the proof of Proposition 5.1, if $q/n \to 0$ then $n^{-1}Z'(\widehat{\Sigma}^{-1}\otimes \ddot{P})Z - n^{-1}F'(\Sigma^{-1}\otimes I_n)F = o_p(1)$. As $n^{-1}\omega'\omega = O(1)$, $n^{-1}\Delta_1 = O(1)$, $n^{-1}\Delta_2 = O(1)$, and $n^{-1}\omega'\ddot{P}\omega = O(1)$, we have $n^{-1}\widehat{\Omega}_{22}^* - n^{-1}\Omega_{22}^* = o_p(1)$. As shown in the proof of Proposition 9 in Liu (2014), if $q/n \to 0$ then $n^{-1/2}\widehat{\mathrm{E}}[V'(\Sigma^{-1}\otimes \ddot{P})u] - n^{-1/2}\mathrm{E}[V'(\Sigma^{-1}\otimes \ddot{P})u] = o_p(1)$. Hence, it only remains to show that $n^{-1}\widehat{D}_2^* - n^{-1}D_2^* = o_p(1)$.

$$D_2 = -\mathrm{E}[\frac{\partial}{\partial \theta'}h_2(\theta_0)] = \begin{bmatrix} \Upsilon_{1,11} & 0 \\ \Upsilon_{1,12} & 0 \\ 0 & \Upsilon_{1,21} \\ 0 & \Upsilon_{1,22} \end{bmatrix} + \begin{bmatrix} \Upsilon_{2,11} & 0 \\ 0 & \Upsilon_{2,12} \\ \Upsilon_{2,21} & 0 \\ 0 & \Upsilon_{2,22} \end{bmatrix},$$

where $\Upsilon_{1,kl} = [\mathrm{E}(V_k'\ddot{\Xi}_1 u_l), \cdots, \mathrm{E}(V_k'\ddot{\Xi}_p u_l)]'$ and $\Upsilon_{2,kl} = [\mathrm{E}(V_l'\ddot{\Xi}_1' u_k), \cdots, \mathrm{E}(V_l'\ddot{\Xi}_p' u_k)]'$ for

$k, l = 1, 2$. With V_1 and V_2 given by (5.3),

$$\mathrm{E}(V_1' A u_1) = \begin{bmatrix} (\sigma_{12} + \phi_{12,0}\sigma_{11})\mathrm{tr}(A' S^{-1}) + (\lambda_{12,0}\sigma_{11} - \lambda_{11,0}\sigma_{12})\mathrm{tr}(A' W S^{-1}) \\ (\sigma_{11} + \phi_{21,0}\sigma_{12})\mathrm{tr}(A' W S^{-1}) + (\lambda_{21,0}\sigma_{12} - \lambda_{22,0}\sigma_{11})\mathrm{tr}(A' W^2 S^{-1}) \\ (\sigma_{12} + \phi_{12,0}\sigma_{11})\mathrm{tr}(A' W S^{-1}) + (\lambda_{12,0}\sigma_{11} - \lambda_{11,0}\sigma_{12})\mathrm{tr}(A' W^2 S^{-1}) \\ 0_{2\times 1} \end{bmatrix}$$

$$\mathrm{E}(V_2' A u_2) = \begin{bmatrix} (\sigma_{12} + \phi_{21,0}\sigma_{22})\mathrm{tr}(A' S^{-1}) + (\lambda_{21,0}\sigma_{22} - \lambda_{22,0}\sigma_{12})\mathrm{tr}(A' W S^{-1}) \\ (\sigma_{12} + \phi_{21,0}\sigma_{22})\mathrm{tr}(A' W S^{-1}) + (\lambda_{21,0}\sigma_{22} - \lambda_{22,0}\sigma_{12})\mathrm{tr}(A' W^2 S^{-1}) \\ (\sigma_{22} + \phi_{12,0}\sigma_{12})\mathrm{tr}(A' W S^{-1}) + (\lambda_{12,0}\sigma_{12} - \lambda_{11,0}\sigma_{22})\mathrm{tr}(A' W^2 S^{-1}) \\ 0_{2\times 1} \end{bmatrix}$$

$$\mathrm{E}(V_1' A u_2) = \begin{bmatrix} (\sigma_{22} + \phi_{12,0}\sigma_{12})\mathrm{tr}(A' S^{-1}) + (\lambda_{12,0}\sigma_{12} - \lambda_{11,0}\sigma_{22})\mathrm{tr}(A' W S^{-1}) \\ (\sigma_{12} + \phi_{21,0}\sigma_{22})\mathrm{tr}(A' W S^{-1}) + (\lambda_{21,0}\sigma_{22} - \lambda_{22,0}\sigma_{12})\mathrm{tr}(A' W^2 S^{-1}) \\ (\sigma_{22} + \phi_{12,0}\sigma_{12})\mathrm{tr}(A' W S^{-1}) + (\lambda_{12,0}\sigma_{12} - \lambda_{11,0}\sigma_{22})\mathrm{tr}(A' W^2 S^{-1}) \\ 0_{2\times 1} \end{bmatrix}$$

$$\mathrm{E}(V_2' A u_1) = \begin{bmatrix} (\sigma_{11} + \phi_{21,0}\sigma_{12})\mathrm{tr}(A' S^{-1}) + (\lambda_{21,0}\sigma_{12} - \lambda_{22,0}\sigma_{11})\mathrm{tr}(A' W S^{-1}) \\ (\sigma_{11} + \phi_{21,0}\sigma_{12})\mathrm{tr}(A' W S^{-1}) + (\lambda_{21,0}\sigma_{12} - \lambda_{22,0}\sigma_{11})\mathrm{tr}(A' W^2 S^{-1}) \\ (\sigma_{12} + \phi_{12,0}\sigma_{11})\mathrm{tr}(A' W S^{-1}) + (\lambda_{12,0}\sigma_{11} - \lambda_{11,0}\sigma_{12})\mathrm{tr}(A' W^2 S^{-1}) \\ 0_{2\times 1} \end{bmatrix}$$

where A is either $\ddot{\Xi}_r$ or $\ddot{\Xi}_r'$. As $n^{-1}\mathrm{tr}(AS^{-1})$, $n^{-1}\mathrm{tr}(AWS^{-1})$ and $n^{-1}\mathrm{tr}(AW^2S^{-1})$ are $O(1)$, we have $n^{-1}\widehat{D}_2 - n^{-1}D_2 = o_p(1)$, where $\widehat{D}_2$ 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 $q/n \to 0$, $n^{-1}\omega'\ddot{P}Z_k = n^{-1}\omega' F_k + n^{-1}\omega'(I_n - \ddot{P})F_k + n^{-1}\omega'\ddot{P}V_k = n^{-1}\omega' F_k + o_p(1) = O_p(1)$. Hence, $n^{-1}\widehat{D}_2^* = n^{-1}\widehat{D}_2 - n^{-1}[(\widehat{\mu}_3'\widehat{\Sigma}^{-1}) \otimes \omega'\ddot{P}]Z = n^{-1}D_2 + o_p(1)$. The desired result follows. $\square$

Table 1: 3SLS and GMM Estimation ($n_g = 10$, $\bar{g} = 30$)

	$\phi_{21,0} = 0.2$	$\lambda_{11,0} = 0.1$	$\lambda_{21,0} = 0.1$	$\beta_{1,0} = 0.6$	$\gamma_{1,0} = 0.6$
$\sigma_{12} = 0.1$					
3SLS-1	.198(.069)[.069]	.096(.046)[.046]	.104(.045)[.045]	.603(.062)[.062]	.603(.057)[.057]
3SLS-2	.222(.069)[.072]	.087(.035)[.037]	.098(.035)[.035]	.598(.060)[.060]	.603(.052)[.052]
BC3SLS	.194(.070)[.070]	.098(.037)[.037]	.104(.036)[.037]	.604(.061)[.061]	.602(.053)[.053]
GMM-1	.201(.069)[.069]	.095(.038)[.039]	.101(.036)[.036]	.602(.062)[.062]	.600(.055)[.055]
GMM-2	.225(.069)[.074]	.088(.032)[.035]	.097(.031)[.031]	.597(.060)[.061]	.600(.052)[.052]
BCGMM	.197(.070)[.070]	.097(.034)[.034]	.103(.032)[.032]	.603(.061)[.061]	.600(.053)[.053]
$\sigma_{12} = 0.5$					
3SLS-1	.199(.069)[.069]	.096(.044)[.045]	.102(.045)[.045]	.605(.058)[.058]	.603(.055)[.055]
3SLS-2	.224(.069)[.073]	.087(.035)[.037]	.095(.035)[.035]	.605(.056)[.056]	.608(.050)[.050]
BC3SLS	.195(.073)[.073]	.099(.036)[.036]	.103(.037)[.037]	.604(.058)[.058]	.602(.052)[.052]
GMM-1	.200(.071)[.071]	.094(.039)[.040]	.101(.039)[.039]	.603(.057)[.057]	.602(.053)[.053]
GMM-2	.226(.070)[.075]	.087(.033)[.035]	.094(.032)[.033]	.605(.056)[.056]	.607(.050)[.050]
BCGMM	.197(.073)[.073]	.097(.035)[.035]	.102(.034)[.034]	.603(.058)[.058]	.601(.052)[.052]
$\sigma_{12} = 0.9$					
3SLS-1	.200(.070)[.070]	.096(.041)[.041]	.102(.044)[.044]	.606(.047)[.047]	.605(.046)[.047]
3SLS-2	.226(.072)[.077]	.087(.033)[.036]	.092(.035)[.036]	.616(.048)[.050]	.616(.047)[.050]
BC3SLS	.198(.075)[.075]	.098(.035)[.035]	.103(.038)[.038]	.605(.049)[.049]	.604(.049)[.049]
GMM-1	.200(.072)[.072]	.093(.039)[.040]	.101(.043)[.043]	.606(.047)[.047]	.605(.047)[.047]
GMM-2	.226(.073)[.078]	.087(.033)[.035]	.092(.035)[.035]	.616(.048)[.050]	.616(.047)[.050]
BCGMM	.198(.075)[.075]	.097(.035)[.035]	.102(.037)[.037]	.605(.049)[.049]	.604(.048)[.049]

Mean(SD)[RMSE]

Table 2: 3SLS and GMM Estimation ($n_g = 15$, $\bar{g} = 30$)

	$\phi_{21,0} = 0.2$	$\lambda_{11,0} = 0.1$	$\lambda_{21,0} = 0.1$	$\beta_{1,0} = 0.6$	$\gamma_{1,0} = 0.6$
$\sigma_{12} = 0.1$					
3SLS-1	.201(.053)[.053]	.099(.034)[.034]	.099(.033)[.033]	.598(.051)[.051]	.603(.046)[.046]
3SLS-2	.217(.054)[.057]	.097(.027)[.028]	.095(.028)[.029]	.596(.050)[.050]	.602(.042)[.042]
BC3SLS	.199(.055)[.055]	.100(.028)[.028]	.100(.029)[.029]	.598(.051)[.051]	.602(.043)[.043]
GMM-1	.204(.052)[.052]	.099(.027)[.027]	.098(.025)[.025]	.597(.051)[.051]	.601(.042)[.042]
GMM-2	.219(.053)[.057]	.096(.024)[.024]	.095(.023)[.024]	.595(.050)[.050]	.600(.041)[.041]
BCGMM	.201(.054)[.054]	.099(.024)[.024]	.099(.024)[.024]	.597(.050)[.051]	.601(.041)[.041]
$\sigma_{12} = 0.5$					
3SLS-1	.202(.053)[.053]	.099(.033)[.033]	.099(.033)[.033]	.599(.047)[.047]	.603(.043)[.043]
3SLS-2	.217(.054)[.057]	.096(.027)[.027]	.094(.028)[.028]	.601(.046)[.046]	.606(.040)[.040]
BC3SLS	.199(.056)[.056]	.100(.028)[.028]	.099(.029)[.029]	.598(.047)[.047]	.602(.041)[.041]
GMM-1	.204(.052)[.052]	.099(.028)[.028]	.098(.027)[.027]	.598(.047)[.047]	.602(.040)[.040]
GMM-2	.219(.053)[.057]	.096(.024)[.024]	.094(.024)[.025]	.601(.046)[.046]	.605(.038)[.039]
BCGMM	.201(.055)[.055]	.100(.025)[.025]	.099(.025)[.025]	.598(.047)[.047]	.601(.040)[.040]
$\sigma_{12} = 0.9$					
3SLS-1	.203(.052)[.052]	.100(.030)[.030]	.098(.032)[.032]	.602(.037)[.037]	.603(.036)[.036]
3SLS-2	.218(.055)[.058]	.095(.026)[.026]	.093(.028)[.028]	.608(.038)[.039]	.610(.037)[.039]
BC3SLS	.200(.057)[.057]	.100(.026)[.026]	.099(.029)[.029]	.601(.039)[.039]	.603(.038)[.038]
GMM-1	.203(.053)[.053]	.099(.028)[.028]	.097(.030)[.030]	.602(.037)[.037]	.603(.036)[.036]
GMM-2	.219(.055)[.058]	.096(.024)[.025]	.094(.026)[.027]	.608(.038)[.039]	.610(.037)[.038]
BCGMM	.201(.057)[.057]	.100(.025)[.025]	.098(.027)[.027]	.601(.039)[.039]	.602(.038)[.038]

Mean(SD)[RMSE]

Table 3: 3SLS and GMM Estimation ($n_g = 15$, $\bar{g} = 60$)

	$\phi_{21,0} = 0.2$	$\lambda_{11,0} = 0.1$	$\lambda_{21,0} = 0.1$	$\beta_{1,0} = 0.6$	$\gamma_{1,0} = 0.6$
$\sigma_{12} = 0.1$					
3SLS-1	.200(.038)[.038]	.099(.023)[.023]	.100(.025)[.025]	.600(.035)[.035]	.600(.032)[.032]
3SLS-2	.215(.038)[.041]	.096(.019)[.020]	.097(.020)[.021]	.598(.035)[.035]	.601(.030)[.030]
BC3SLS	.198(.039)[.039]	.099(.020)[.020]	.101(.021)[.021]	.600(.035)[.035]	.601(.031)[.031]
GMM-1	.201(.037)[.037]	.099(.019)[.019]	.100(.019)[.019]	.600(.035)[.035]	.600(.030)[.030]
GMM-2	.216(.038)[.041]	.096(.017)[.017]	.097(.018)[.018]	.598(.035)[.035]	.600(.029)[.029]
BCGMM	.200(.038)[.038]	.099(.017)[.017]	.101(.018)[.018]	.600(.035)[.035]	.600(.030)[.030]
$\sigma_{12} = 0.5$					
3SLS-1	.201(.038)[.038]	.099(.022)[.022]	.100(.024)[.024]	.600(.033)[.033]	.601(.030)[.030]
3SLS-2	.215(.039)[.041]	.095(.019)[.019]	.096(.020)[.021]	.602(.032)[.032]	.605(.028)[.029]
BC3SLS	.199(.040)[.040]	.099(.019)[.019]	.101(.021)[.021]	.599(.033)[.033]	.601(.029)[.029]
GMM-1	.201(.038)[.038]	.098(.019)[.019]	.100(.021)[.021]	.600(.033)[.033]	.601(.029)[.029]
GMM-2	.216(.038)[.041]	.095(.017)[.018]	.097(.018)[.019]	.602(.032)[.032]	.604(.028)[.028]
BCGMM	.199(.040)[.040]	.099(.017)[.017]	.101(.019)[.019]	.599(.033)[.033]	.601(.029)[.029]
$\sigma_{12} = 0.9$					
3SLS-1	.201(.038)[.038]	.099(.021)[.021]	.100(.024)[.024]	.601(.027)[.027]	.602(.026)[.026]
3SLS-2	.216(.040)[.044]	.094(.018)[.019]	.095(.020)[.021]	.607(.028)[.029]	.608(.026)[.027]
BC3SLS	.200(.042)[.042]	.099(.018)[.018]	.101(.021)[.021]	.600(.029)[.029]	.601(.027)[.027]
GMM-1	.201(.039)[.039]	.098(.020)[.020]	.100(.023)[.023]	.601(.028)[.028]	.602(.026)[.026]
GMM-2	.216(.041)[.044]	.095(.017)[.018]	.096(.020)[.020]	.607(.028)[.029]	.608(.026)[.027]
BCGMM	.200(.042)[.042]	.099(.018)[.018]	.100(.021)[.021]	.600(.028)[.028]	.601(.027)[.027]

Mean(SD)[RMSE]

Table 4: 3SLS and GMM Estimation ($n_g = 10$, $\bar{g} = 30$)

	$\phi_{21,0} = 0.2$	$\lambda_{11,0} = 0.1$	$\lambda_{21,0} = 0.1$	$\beta_{1,0} = 0.3$	$\gamma_{1,0} = 0.3$
$\sigma_{12} = 0.1$					
3SLS-1	.195(.143)[.143]	.089(.100)[.101]	.108(.097)[.097]	.304(.065)[.065]	.305(.060)[.060]
3SLS-2	.267(.149)[.164]	.076(.051)[.056]	.093(.052)[.053]	.294(.061)[.061]	.298(.052)[.052]
BC3SLS	.181(.153)[.154]	.101(.059)[.059]	.105(.057)[.057]	.307(.063)[.063]	.302(.053)[.053]
GMM-1	.212(.144)[.144]	.093(.059)[.060]	.099(.053)[.053]	.302(.062)[.062]	.299(.054)[.054]
GMM-2	.275(.151)[.169]	.078(.044)[.049]	.089(.042)[.043]	.293(.061)[.061]	.294(.052)[.053]
BCGMM	.192(.150)[.150]	.099(.048)[.048]	.103(.046)[.046]	.305(.062)[.062]	.300(.052)[.052]
$\sigma_{12} = 0.5$					
3SLS-1	.201(.147)[.147]	.091(.096)[.097]	.103(.095)[.095]	.307(.061)[.061]	.305(.057)[.058]
3SLS-2	.276(.150)[.168]	.076(.051)[.056]	.087(.052)[.053]	.305(.056)[.057]	.307(.048)[.049]
BC3SLS	.200(.321)[.321]	.107(.199)[.199]	.105(.070)[.070]	.310(.102)[.103]	.304(.074)[.074]
GMM-1	.208(.149)[.149]	.091(.065)[.065]	.100(.061)[.061]	.305(.058)[.058]	.303(.052)[.052]
GMM-2	.281(.152)[.172]	.077(.045)[.050]	.086(.044)[.046]	.305(.056)[.057]	.305(.048)[.049]
BCGMM	.198(.160)[.160]	.099(.052)[.052]	.103(.053)[.053]	.306(.060)[.060]	.302(.053)[.053]
$\sigma_{12} = 0.9$					
3SLS-1	.205(.149)[.149]	.088(.087)[.088]	.100(.092)[.092]	.310(.049)[.050]	.309(.048)[.049]
3SLS-2	.283(.154)[.175]	.076(.050)[.055]	.081(.052)[.055]	.324(.050)[.055]	.324(.048)[.053]
BC3SLS	.205(.193)[.193]	.092(.250)[.250]	.107(.100)[.101]	.311(.064)[.065]	.310(.053)[.054]
GMM-1	.203(.158)[.158]	.088(.075)[.076]	.102(.080)[.080]	.309(.049)[.050]	.309(.048)[.049]
GMM-2	.285(.155)[.177]	.077(.048)[.053]	.083(.050)[.052]	.324(.050)[.055]	.324(.047)[.053]
BCGMM	.207(.159)[.159]	.092(.157)[.158]	.105(.082)[.082]	.311(.052)[.053]	.310(.049)[.050]

Mean(SD)[RMSE]

Table 5: 3SLS and GMM Estimation ($n_g = 15$, $\bar{g} = 30$)

	$\phi_{21,0} = 0.2$	$\lambda_{11,0} = 0.1$	$\lambda_{21,0} = 0.1$	$\beta_{1,0} = 0.3$	$\gamma_{1,0} = 0.3$
$\sigma_{12} = 0.1$					
3SLS-1	.203(.108)[.108]	.097(.070)[.070]	.100(.068)[.068]	.299(.052)[.052]	.304(.046)[.047]
3SLS-2	.251(.113)[.124]	.092(.040)[.041]	.088(.044)[.046]	.294(.050)[.050]	.299(.040)[.040]
BC3SLS	.193(.115)[.115]	.101(.043)[.043]	.100(.046)[.046]	.299(.051)[.051]	.302(.041)[.041]
GMM-1	.210(.104)[.104]	.098(.040)[.040]	.097(.037)[.037]	.298(.051)[.051]	.300(.040)[.040]
GMM-2	.256(.112)[.126]	.091(.032)[.033]	.089(.033)[.035]	.294(.050)[.050]	.298(.039)[.039]
BCGMM	.200(.114)[.114]	.100(.033)[.033]	.100(.034)[.034]	.298(.050)[.050]	.301(.039)[.039]
$\sigma_{12} = 0.5$					
3SLS-1	.205(.107)[.107]	.097(.068)[.068]	.097(.067)[.067]	.300(.048)[.048]	.305(.044)[.044]
3SLS-2	.254(.113)[.125]	.091(.040)[.041]	.087(.043)[.046]	.302(.046)[.046]	.306(.037)[.038]
BC3SLS	.195(.119)[.119]	.102(.043)[.043]	.100(.047)[.047]	.300(.048)[.048]	.303(.040)[.040]
GMM-1	.210(.104)[.104]	.098(.043)[.043]	.096(.041)[.042]	.300(.047)[.047]	.302(.038)[.038]
GMM-2	.257(.112)[.126]	.091(.033)[.034]	.088(.035)[.037]	.302(.046)[.046]	.305(.036)[.036]
BCGMM	.200(.117)[.117]	.100(.035)[.035]	.099(.038)[.038]	.299(.047)[.047]	.302(.038)[.038]
$\sigma_{12} = 0.9$					
3SLS-1	.208(.105)[.106]	.096(.063)[.063]	.095(.066)[.066]	.305(.038)[.038]	.306(.037)[.037]
3SLS-2	.257(.114)[.127]	.089(.039)[.040]	.086(.043)[.045]	.314(.039)[.041]	.316(.037)[.040]
BC3SLS	.201(.118)[.118]	.102(.047)[.047]	.099(.047)[.047]	.304(.039)[.040]	.305(.037)[.038]
GMM-1	.208(.110)[.111]	.098(.049)[.049]	.096(.055)[.055]	.304(.038)[.038]	.305(.036)[.036]
GMM-2	.258(.114)[.128]	.090(.035)[.037]	.087(.039)[.041]	.315(.039)[.041]	.316(.036)[.040]
BCGMM	.203(.117)[.117]	.101(.039)[.039]	.098(.043)[.043]	.304(.039)[.039]	.305(.037)[.037]

Mean(SD)[RMSE]

Table 6: 3SLS and GMM Estimation ($n_g = 15$, $\bar{g} = 60$)

	$\phi_{21,0} = 0.2$	$\lambda_{11,0} = 0.1$	$\lambda_{21,0} = 0.1$	$\beta_{1,0} = 0.3$	$\gamma_{1,0} = 0.3$
$\sigma_{12} = 0.1$					
3SLS-1	.201(.077)[.077]	.098(.047)[.047]	.101(.050)[.050]	.300(.036)[.036]	.301(.032)[.032]
3SLS-2	.246(.080)[.092]	.090(.028)[.030]	.092(.031)[.032]	.297(.035)[.035]	.299(.029)[.029]
BC3SLS	.195(.080)[.080]	.099(.029)[.029]	.102(.032)[.032]	.301(.035)[.035]	.301(.029)[.029]
GMM-1	.205(.073)[.074]	.098(.028)[.028]	.100(.027)[.027]	.300(.035)[.035]	.299(.029)[.029]
GMM-2	.249(.079)[.093]	.090(.022)[.025]	.093(.024)[.025]	.297(.035)[.035]	.298(.028)[.028]
BCGMM	.198(.079)[.079]	.099(.023)[.023]	.102(.024)[.024]	.300(.035)[.035]	.300(.028)[.028]
$\sigma_{12} = 0.5$					
3SLS-1	.202(.077)[.077]	.097(.045)[.045]	.101(.049)[.049]	.301(.033)[.033]	.302(.030)[.030]
3SLS-2	.247(.081)[.094]	.089(.028)[.030]	.091(.030)[.032]	.303(.032)[.032]	.306(.026)[.027]
BC3SLS	.196(.084)[.084]	.099(.029)[.030]	.102(.033)[.033]	.300(.033)[.033]	.301(.028)[.028]
GMM-1	.203(.075)[.075]	.097(.030)[.030]	.100(.031)[.031]	.300(.033)[.033]	.301(.028)[.028]
GMM-2	.249(.080)[.094]	.089(.024)[.026]	.092(.025)[.027]	.303(.032)[.032]	.305(.026)[.026]
BCGMM	.198(.083)[.083]	.099(.025)[.025]	.102(.027)[.027]	.300(.033)[.033]	.301(.028)[.028]
$\sigma_{12} = 0.9$					
3SLS-1	.203(.077)[.077]	.097(.042)[.042]	.100(.048)[.048]	.302(.027)[.027]	.303(.026)[.026]
3SLS-2	.250(.083)[.097]	.087(.028)[.030]	.090(.031)[.032]	.311(.028)[.030]	.313(.026)[.029]
BC3SLS	.199(.086)[.086]	.099(.029)[.029]	.101(.033)[.033]	.301(.028)[.028]	.302(.027)[.027]
GMM-1	.201(.081)[.081]	.096(.036)[.036]	.101(.042)[.042]	.302(.028)[.028]	.303(.026)[.026]
GMM-2	.250(.083)[.097]	.089(.026)[.028]	.091(.029)[.030]	.311(.028)[.030]	.313(.026)[.029]
BCGMM	.200(.085)[.085]	.098(.028)[.028]	.101(.031)[.031]	.301(.028)[.028]	.302(.026)[.027]

Mean(SD)[RMSE]