Identification and Efficient Estimation of Simultaneous Equations Network Models*

Xiaodong Liu†

Department of Economics, University of Colorado, Boulder, CO 80309, USA

March, 2013

Abstract

We consider identification and estimation of social network models in a system of simultaneous equations. We show that, with or without row-normalization of the social adjacency matrix, the network model has different equilibrium implications, needs different identification conditions, and requires different estimation strategies. When the adjacency matrix is not row-normalized, different positions of the agents in a network captured by the Bonacich centrality can be used to identify social interaction effects and improve estimation efficiency. We show that the identification condition for the model with a non-row-normalized adjacency matrix is weaker than that for the model with a row-normalized adjacency matrix. We suggest 2SLS and 3SLS estimators using instruments based on the Bonacich centrality of each network to improve estimation efficiency. The number of such instruments depends on the number of networks. When there are many networks in the data, the proposed estimators may have an asymptotic bias due to the presence of many instruments. We propose a bias-correction procedure for the many-instrument bias. Simulation experiments show that the bias-corrected estimators perform well in finite samples.

JEL classification: C31, C36

Key words: social networks, simultaneous equations, identification, efficient estimation

*I am thankful to the seminar participants at Syracuse University and University of Colorado for helpful comments and suggestions.

†Phone: 1(303)4927414. Fax: 1(303)4928960. Email: xiaodong.liu@colorado.edu.

1 Introduction

Since the seminal work by Manski (1993), research on social network models has attracted a lot of attention (see Blume et al., 2011, for a recent survey). In a social network model, agents interact with each other through network connections, which are captured by a social adjacency matrix. According to Manski (1993), an agent' choice or outcome may be influenced by peer choices or outcomes (*the endogenous effect*), by peer exogenous characteristics (*the contextual effect*), and/or by the common environment of the network (*the correlated effect*). It is the main interest of social network research to separately identify and estimate different social interaction effects.

Manski (1993) considers a linear-in-means model, where the endogenous effect is based on the rational expectation of the outcomes of all agents in the network. Manski shows that the linear-in-means specification suffers from the "reflection problem" so that endogenous and contextual effects cannot be separately identified. Lee (2007) introduces a model with multiple networks where an agent is equally influenced by all the other agents in the same network. Lee's social network model can be identified using the variation in network sizes. The identification, however, can be weak if all of networks are large. Bramoullé et al. (2009) generalize Lee's social network model to a general *local-average* model, where endogenous and contextual effects are represented, respectively, by the *average* outcome and *average* characteristics of an agent's connections (or friends). Based on the important observation that in a social network, an agent's friend's friend may not be a friend of that agent, Bramoullé et al. (2009) use the intransitivity in network connections as an exclusion restriction to identify different social interaction effects.

Liu and Lee (2010) studies the efficient estimation of the *local-aggregate* social network model, where the endogenous effect is given by the *aggregate* outcome of an agent's friends. They show that, for the local-aggregate model, different positions of the agents in a network captured by the Bonacich (1987) centrality can be used as additional instruments to improve estimation efficiency. Liu et al. (2012) give the identification condition for the local-aggregate model and show that the condition is weaker than that for the local-average model derived by Bramoullé et al. (2009). Liu et al. (2012) also propose a J test for the specification of network models.

The above mentioned papers focus on single-equation network models with only one activity. However, in real life, an agent's decision usually involves more than one activity. For example, a student may need to balance time between study and extracurriculars and a firm may need to allocate resources between production and RND. In a recent paper, Cohen-Cole et al. (2012) consider the identification and estimation of *local-average* network models in the framework of simultaneous equations. Besides endogenous, contextual, and correlated effects as in single-equation network models, the simultaneous equations network model also incorporates *the simultaneity effect*, where an agent's outcome in a certain activity may depend on his/her outcome in a related activity, and *the cross-activity peer effect*, where an agent's outcome in a certain activity may depend on peer outcomes in a related activity. Cohen-Cole et al. (2012) derive the identification conditions for the various social interaction effects and generalize the spatial 2SLS and 3SLS estimators in Kelejian and Prucha (2004) to estimate the simultaneous equations network model.

In this paper, we consider the identification and efficient estimation of the *local-aggregate* network model in a system of simultaneous equations. We show that, similar to the single-equation network model, the Bonacich centrality provides additional information to achieve model identification and to improve estimation efficiency. We derive the identification condition for the local-aggregate simultaneous equations network model, and show that the condition is weaker than that for the local-average model. For efficient estimation, we suggest 2SLS and 3SLS estimators using instruments based on the Bonacich centrality of each network. As the number of such instruments depends on the number of networks, the 2SLS and 3SLS estimators would have an asymptotic many-instrument bias (Bekker, 1994) when there are many networks in the data. Hence, we propose a bias-correction procedure based on the estimated leading-order term of the asymptotic bias. Monte Carlo experiments show that the bias-corrected estimators perform well in finite samples.

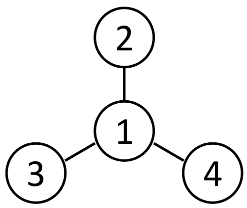
The rest of the paper is organized as follows. Section 2 introduces a network game which motivates the specification of the econometric model presented in Section 3. Section 4 derives the identification conditions and Section 5 proposes 2SLS and 3SLS estimators for the model. The regularity assumptions and detailed proofs are given in the Appendix. Monte Carlo evidence on the finite sample performance of the proposed estimators is given in Section 6. Section 7 briefly concludes.

2 Theoretical Model

2.1 The network game

Suppose there is a finite set of agents $N = \{1, ..., n\}$ in a network. We keep track of social connections in the network through its adjacency matrix $G = [g_{ij}]$, where $g_{ij} = 1$ if i and j are friends and $g_{ij} = 0$

otherwise.[1] Let $G^* = [g^*_{ij}]$, with $g^*_{ij} = g_{ij}/\sum_{j=1}^n g_{ij}$, denote the row-normalized adjacency matrix such that each row of G^* adds up to one. Figure 1 gives an example of G and G^* for a star-shaped network.



$$G = \begin{bmatrix} 0 & 1 & 1 & 1 \\ 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \end{bmatrix} \qquad G^* = \begin{bmatrix} 0 & 1/3 & 1/3 & 1/3 \\ 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \end{bmatrix}$$

Figure 1: an example of G and G^* for a star-shaped network.

Given the network structure represented by G, agent i chooses y_{1i} and y_{2i}, the respective efforts of two related activities, to maximize the following linear quadratic utility function

$$\begin{aligned} u(y_{1i}, y_{2i}) &= \pi^*_{1i} y_{1i} + \pi^*_{2i} y_{2i} - \frac{1}{2}\phi^*_1 y^2_{1i} - \frac{1}{2}\phi^*_2 y^2_{2i} + \phi^* y_{1i} y_{2i} \qquad (1) \\ &\quad + \lambda^*_{11} \textstyle\sum_{j=1}^n g_{ij} y_{1i} y_{1j} + \lambda^*_{22} \sum_{j=1}^n g_{ij} y_{2i} y_{2j} + \lambda^*_{21} \sum_{j=1}^n g_{ij} y_{1i} y_{2j} + \lambda^*_{12} \sum_{j=1}^n g_{ij} y_{2i} y_{1j}. \end{aligned}$$

As in the standard linear-quadratic utility for a single activity model (Ballester et al., 2006), π^*_{1i} and π^*_{2i} capture *ex ante* individual heterogeneity. The cross-effects between own efforts for different activities are given by

$$\frac{\partial^2 u(y_{1i}, y_{2i})}{\partial y_{1i} \partial y_{2i}} = \frac{\partial^2 u(y_{1i}, y_{2i})}{\partial y_{2i} \partial y_{1i}} = \phi^*.$$

The cross-effects between own and peer efforts for the same activity are

$$\frac{\partial^2 u(y_{1i}, y_{2i})}{\partial y_{1i} \partial y_{1j}} = \lambda^*_{11} g_{ij} \text{ and } \frac{\partial^2 u(y_{1i}, y_{2i})}{\partial y_{2i} \partial y_{2j}} = \lambda^*_{22} g_{ij},$$

which may indicate strategic substitutability or complementarity depending on the signs of λ^*_{11} and λ^*_{22}. The cross-effects between own and peer efforts for different activities are given by

$$\frac{\partial u(y_{1i}, y_{2i})}{\partial y_{1i} \partial y_{2j}} = \lambda^*_{21} g_{ij} \text{ and } \frac{\partial u(y_{1i}, y_{2i})}{\partial y_{2i} \partial y_{1j}} = \lambda^*_{12} g_{ij},$$

which may indicate strategic substitutability or complementarity depending on the signs of λ^*_{21} and

[1] For ease of presentation, we focus on the case where the connections are undirected and no agent is isolated so that G is symmetric and $\sum_{j=1}^n g_{ij} \neq 0$ for all i. The identification result and estimation method of the paper hold for a directed network with an asymmetric G.

λ_{12}^*.

From the first order conditions of utility maximization, we have the *best-response functions* as

$$y_{1i} = \phi_1 y_{2i} + \lambda_{11} \sum_{j=1}^n g_{ij} y_{1j} + \lambda_{21} \sum_{j=1}^n g_{ij} y_{2j} + \pi_{1i}, \quad (2)$$

$$y_{2i} = \phi_2 y_{1i} + \lambda_{22} \sum_{j=1}^n g_{ij} y_{2j} + \lambda_{12} \sum_{j=1}^n g_{ij} y_{1j} + \pi_{2i}, \quad (3)$$

where $\phi_1 = \phi^*/\phi_1^*$, $\phi_2 = \phi^*/\phi_2^*$, $\lambda_{11} = \lambda_{11}^*/\phi_1^*$, $\lambda_{22} = \lambda_{22}^*/\phi_2^*$, $\lambda_{21} = \lambda_{21}^*/\phi_1^*$, $\lambda_{12} = \lambda_{12}^*/\phi_2^*$, $\pi_{1i} = \pi_{1i}^*/\phi_1^*$, $\pi_{2i} = \pi_{2i}^*/\phi_2^*$. In (2) and (3), agent i's best-response effort of a certain activity depends on the *aggregate* efforts of his/her friends of that activity and a related activity. Therefore, we call this model the *local-aggregate* network game. In matrix form, the best-response functions are

$$Y_1 = \phi_1 Y_2 + \lambda_{11} G Y_1 + \lambda_{21} G Y_2 + \Pi_1, \quad (4)$$

$$Y_2 = \phi_2 Y_1 + \lambda_{22} G Y_2 + \lambda_{12} G Y_1 + \Pi_2, \quad (5)$$

where $Y_k = (y_{k1}, \cdots, y_{kn})'$ and $\Pi_k = (\pi_{k1}, \cdots, \pi_{kn})'$ for $k = 1, 2$.

The reduced-form equations of (4) and (5) are

$$SY_1 = (I - \lambda_{22} G)\Pi_1 + (\phi_1 I + \lambda_{21} G)\Pi_2,$$

$$SY_2 = (I - \lambda_{11} G)\Pi_2 + (\phi_2 I + \lambda_{12} G)\Pi_1,$$

where I is a conformable identity matrix and

$$S = (1 - \phi_1\phi_2)I - (\lambda_{11} + \lambda_{22} + \phi_1\lambda_{12} + \phi_2\lambda_{21})G + (\lambda_{11}\lambda_{22} - \lambda_{12}\lambda_{21})G^2. \quad (6)$$

If S is nonsingular[2], then the local-aggregate network game has a unique Nash equilibrium in pure strategies with the equilibrium efforts given by

$$Y_1^* = S^{-1}[(I - \lambda_{22} G)\Pi_1 + (\phi_1 I + \lambda_{21} G)\Pi_2], \quad (7)$$

$$Y_2^* = S^{-1}[(I - \lambda_{11} G)\Pi_2 + (\phi_2 I + \lambda_{12} G)\Pi_1]. \quad (8)$$

[2] A sufficient condition for the nonsingularity of S is $|\phi_1\phi_2| + |\lambda_{11} + \lambda_{22} + \phi_1\lambda_{12} + \phi_2\lambda_{21}| \cdot ||G||_\infty + |\lambda_{11}\lambda_{22} - \lambda_{12}\lambda_{21}| \cdot ||G||_\infty^2 < 1$, where $||\cdot||_\infty$ is the row-sum matrix norm.

2.2 Local aggregate versus local average: equilibrium comparison

In a recent paper, Cohen-Cole et al. (2012) consider a network game with utility function

$$\begin{aligned} u(y_{1i}, y_{2i}) &= \pi_{1i}^* y_{1i} + \pi_{2i}^* y_{2i} - \frac{1}{2}\phi_1^* y_{1i}^2 - \frac{1}{2}\phi_2^* y_{2i}^2 + \phi^* y_{1i} y_{2i} \\ &\quad + \lambda_{11}^* \textstyle\sum_{j=1}^n g_{ij}^* y_{1i} y_{1j} + \lambda_{22}^* \sum_{j=1}^n g_{ij}^* y_{2i} y_{2j} + \lambda_{21}^* \sum_{j=1}^n g_{ij}^* y_{1i} y_{2j} + \lambda_{12}^* \sum_{j=1}^n g_{ij}^* y_{2i} y_{1j}. \end{aligned} \tag{9}$$

From the first order conditions of maximizing (9), the best-response functions of the network game are

$$\begin{aligned} y_{1i} &= \phi_1 y_{2i} + \lambda_{11} \textstyle\sum_{j=1}^n g_{ij}^* y_{1j} + \lambda_{21} \sum_{j=1}^n g_{ij}^* y_{2j} + \pi_{1i}, \\ y_{2i} &= \phi_2 y_{1i} + \lambda_{22} \textstyle\sum_{j=1}^n g_{ij}^* y_{2j} + \lambda_{12} \sum_{j=1}^n g_{ij}^* y_{1j} + \pi_{2i}, \end{aligned}$$

or, in matrix form,

$$Y_1 = \phi_1 Y_2 + \lambda_{11} G^* Y_1 + \lambda_{21} G^* Y_2 + \Pi_1, \tag{10}$$

$$Y_2 = \phi_2 Y_1 + \lambda_{22} G^* Y_2 + \lambda_{12} G^* Y_1 + \Pi_2. \tag{11}$$

As G^* is row-normalized, in (10) and (11), agent i's best-response effort of a certain activity depends on the *average* efforts of his/her friends of that activity and a related activity. Therefore, we call this model the *local-average* network game. Cohen-Cole et al. (2012) show that, if S^* is nonsingular, where

$$S^* = (1 - \phi_1\phi_2)I - (\lambda_{11} + \lambda_{22} + \phi_1\lambda_{12} + \phi_2\lambda_{21})G^* + (\lambda_{11}\lambda_{22} - \lambda_{12}\lambda_{21})G^{*2},$$

then the network game with payoffs (9) has a unique Nash equilibrium in pure strategies given by

$$Y_1^* = S^{*-1}[(I - \lambda_{22}G^*)\Pi_1 + (\phi_1 I + \lambda_{21}G^*)\Pi_2], \tag{12}$$

$$Y_2^* = S^{*-1}[(I - \lambda_{11}G^*)\Pi_2 + (\phi_2 I + \lambda_{12}G^*)\Pi_1]. \tag{13}$$

Although the best-response functions of the local-aggregate and local-average network games share similar functional forms, they have different implications. As pointed out by Liu et al. (2012), in the *local-aggregate* game, even if agents are ex ante identical in terms of individual attributes Π_1 and Π_2, agents with different positions in the network would have different equilibrium payoffs. On the other hand, the *local-average* game is based on the mechanism of social conformism.[3] As the positions in the network do not matter in the local-average game, the equilibrium efforts and payoffs would be the same if all agents are ex ante identical.

To illustrate this point, suppose the agents in a network are ex ante identical such that $\Pi_1 = \pi_1 l_n$ and $\Pi_2 = \pi_2 l_n$, where π_1, π_2 are constant scalars and l_n is an $n \times 1$ vector of ones. As $G^* l_n = G^{*2} l_n = l_n$, it follows from (12) and (13) that $Y_1^* = c_1 l_n$ and $Y_2^* = c_2 l_n$, where

$$\begin{aligned} c_1 &= [(1-\lambda_{22})\pi_1 + (\phi_1 + \lambda_{21})\pi_2]/[(1-\phi_1\phi_2) - (\lambda_{11} + \lambda_{22} + \phi_1\lambda_{12} + \phi_2\lambda_{21}) + (\lambda_{11}\lambda_{22} - \lambda_{12}\lambda_{21})], \\ c_2 &= [(1-\lambda_{11})\pi_2 + (\phi_2 + \lambda_{12})\pi_1]/[(1-\phi_1\phi_2) - (\lambda_{11} + \lambda_{22} + \phi_1\lambda_{12} + \phi_2\lambda_{21}) + (\lambda_{11}\lambda_{22} - \lambda_{12}\lambda_{21})]. \end{aligned}$$

Thus, for the local-average network game, the equilibrium efforts and payoffs are the same for all agents. On the other hand, for the local-aggregate network game, it follows from (7) and (8) that

$$\begin{aligned} Y_1^* &= S^{-1}[(I - \lambda_{22}G)\pi_1 + (\phi_1 I + \lambda_{21}G)\pi_2]l_n, \\ Y_2^* &= S^{-1}[(I - \lambda_{11}G)\pi_2 + (\phi_2 I + \lambda_{12}G)\pi_1]l_n. \end{aligned}$$

Thus, the agents would have different equilibrium efforts and payoffs if Gl_n is not proportional to l_n.[4]

Therefore, the local-aggregate and local-average network games have different equilibrium and policy implications. The following sections show that the econometric model for the local-aggregate network game has some interesting features that requires different identification conditions and estimation methods from those for the local-average model studied by Cohen-Cole et al. (2012).

[3] Liu et al. (2012) show that the best-response function of the local-average network game can be derived from a setting where an agent will be punished if he deviates from the "social norm" (the average behavior of his friends). Therefore, if the agents are identical ex ante, they would behave the same in equilibrium.

[4] If $Gl_n = cl_n$ for some constant scalar c, i.e., all agents have the same number of friends, then it follows from (7) and (8) that $Y_1^* = c_3 l_n$ and $Y_2^* = c_4 l_n$, where

$$\begin{aligned} c_3 &= [(1-\lambda_{22}c)\pi_1 + (\phi_1 + \lambda_{21}c)\pi_2]/[(1-\phi_1\phi_2) - (\lambda_{11} + \lambda_{22} + \phi_1\lambda_{12} + \phi_2\lambda_{21})c + (\lambda_{11}\lambda_{22} - \lambda_{12}\lambda_{21})c^2], \\ c_4 &= [(1-\lambda_{11}c)\pi_2 + (\phi_2 + \lambda_{12}c)\pi_1]/[(1-\phi_1\phi_2) - (\lambda_{11} + \lambda_{22} + \phi_1\lambda_{12} + \phi_2\lambda_{21})c + (\lambda_{11}\lambda_{22} - \lambda_{12}\lambda_{21})c^2]. \end{aligned}$$

Thus, all agents have the same equilibrium payoffs according to (1).

3 Econometric Model

3.1 The local-aggregate simultaneous equations network model

The econometric network model follows the best-response functions (4) and (5). Suppose the n observations in the data are partitioned into $\bar{r}$ networks, with n_r agents in the rth network. For the rth network, let

$$\begin{aligned} \Pi_{1,r} &= X_r\beta_1 + G_rX_r\gamma_1 + \alpha_{1,r}l_{n_r} + \epsilon_{1,r}, \\ \Pi_{2,r} &= X_r\beta_2 + G_rX_r\gamma_2 + \alpha_{2,r}l_{n_r} + \epsilon_{2,r}. \end{aligned}$$

where X_r is an $n_r \times k_x$ matrix of exogenous variables, G_r is the adjacency matrix of network r, l_{n_r} is an $n_r \times 1$ vector of ones, and $\epsilon_{1,r}, \epsilon_{2,r}$ are $n_r \times 1$ vectors of disturbances. It follows by (4) and (5) that $Y_{1,r}$ and $Y_{2,r}$, which are $n_r \times 1$ vectors of observed choices/outcomes of two related activities for the agents in the rth network, are given by

$$\begin{aligned} Y_{1,r} &= \phi_1Y_{2,r} + \lambda_{11}G_rY_{1,r} + \lambda_{21}G_rY_{2,r} + X_r\beta_1 + G_rX_r\gamma_1 + \alpha_{1,r}l_{n_r} + \epsilon_{1,r}, \\ Y_{2,r} &= \phi_2Y_{1,r} + \lambda_{22}G_rY_{2,r} + \lambda_{12}G_rY_{1,r} + X_r\beta_2 + G_rX_r\gamma_2 + \alpha_{2,r}l_{n_r} + \epsilon_{2,r}. \end{aligned}$$

Let diag$\{A_s\}$ denote a "generalized" block diagonal matrix with diagonal blocks being $n_s \times m_s$ matrices A_s's. For $k = 1, 2$, let $Y_k = (Y'_{k,1}, \cdots, Y'_{k,\bar{r}})'$, $X = (X'_1, \cdots, X'_{\bar{r}})'$, $\alpha_k = (\alpha_{k,1}, \cdots, \alpha_{k,\bar{r}})'$, $\epsilon_k = (\epsilon'_{k,1}, \cdots, \epsilon'_{k,\bar{r}})'$, $L = \text{diag}\{l_{n_r}\}_{r=1}^{\bar{r}}$ and $G = \text{diag}\{G_r\}_{r=1}^{\bar{r}}$. Then, for all the $\bar{r}$ networks,

$$Y_1 = \phi_1Y_2 + \lambda_{11}GY_1 + \lambda_{21}GY_2 + X\beta_1 + GX\gamma_1 + L\alpha_1 + \epsilon_1, \quad (14)$$

$$Y_2 = \phi_2Y_1 + \lambda_{22}GY_2 + \lambda_{12}GY_1 + X\beta_2 + GX\gamma_2 + L\alpha_2 + \epsilon_2. \quad (15)$$

For $\epsilon_1 = (\epsilon_{11}, \cdots, \epsilon_{1n})'$ and $\epsilon_2 = (\epsilon_{21}, \cdots, \epsilon_{2n})'$, we assume $\text{E}(\epsilon_{1i}) = \text{E}(\epsilon_{2i}) = 0$, $\text{E}(\epsilon_{1i}^2) = \sigma_1^2$, and $\text{E}(\epsilon_{2i}^2) = \sigma_2^2$. Furthermore, we allow the disturbances of the same agent to be correlated across equations by assuming $\text{E}(\epsilon_{1i}\epsilon_{2i}) = \sigma_{12}$ and $\text{E}(\epsilon_{1i}\epsilon_{2j}) = 0$ for $i \neq j$. When S given by (6) is nonsingular, the reduced-form equations of the model are

$$\begin{aligned} Y_1 &= S^{-1}[X(\phi_1\beta_2+\beta_1)+GX(\lambda_{21}\beta_2-\lambda_{22}\beta_1+\phi_1\gamma_2+\gamma_1)+G^2X(\lambda_{21}\gamma_2-\lambda_{22}\gamma_1) \\ &\quad +L(\phi_1\alpha_2+\alpha_1)+GL(\lambda_{21}\alpha_2-\lambda_{22}\alpha_1)]+S^{-1}u_1, \end{aligned} \tag{16}$$

$$\begin{aligned} Y_2 &= S^{-1}[X(\phi_2\beta_1+\beta_2)+GX(\lambda_{12}\beta_1-\lambda_{11}\beta_2+\phi_2\gamma_1+\gamma_2)+G^2X(\lambda_{12}\gamma_1-\lambda_{11}\gamma_2) \\ &\quad +L(\phi_2\alpha_1+\alpha_2)+GL(\lambda_{12}\alpha_1-\lambda_{11}\alpha_2)]+S^{-1}u_2, \end{aligned} \tag{17}$$

where

$$u_1 = (I-\lambda_{22}G)\epsilon_1+(\phi_1 I+\lambda_{21}G)\epsilon_2, \tag{18}$$

$$u_2 = (I-\lambda_{11}G)\epsilon_2+(\phi_2 I+\lambda_{12}G)\epsilon_1. \tag{19}$$

In this model, we allow network-specific effects $\alpha_{1,r}$ and $\alpha_{2,r}$ to depend on X and G by treating α_1 and α_2 as $\bar{r}\times 1$ vectors of unknown parameters (as in a fixed effect panel data model). When the number of network $\bar{r}$ is large, we may have the "incidental parameter" problem (Neyman and Scott, 1948). To avoid this problem, we transform (14) and (15) using a deviation from group mean projector $J=\text{diag}\{J_r\}_{r=1}^{\bar{r}}$ where $J_r = I_{n_r}-\frac{1}{n_r}l_{n_r}l'_{n_r}$. This transformation is analogous to the "within" transformation for fixed effect panel data models. As $JL=0$, the transformed equations are

$$JY_1 = \phi_1 JY_2+\lambda_{11}JGY_1+\lambda_{21}JGY_2+JX\beta_1+JGX\gamma_1+J\epsilon_1, \tag{20}$$

$$JY_2 = \phi_2 JY_1+\lambda_{22}JGY_2+\lambda_{12}JGY_1+JX\beta_2+JGX\gamma_2+J\epsilon_2. \tag{21}$$

Our identification results and estimation methods are based on the transformed model.

3.2 Identification challenges

Analogous to the local-average simultaneous equations network model studied by Cohen-Cole et al. (2012), the local-aggregate simultaneous equations network model given by (14) and (15) incorporates (within-activity) endogenous effects, contextual effects, simultaneity effects, cross-activity peer effects, network correlated effects and cross-activity correlated effects. It is the main purpose of this paper to establish identification conditions and propose efficient estimation methods for the various social interaction effects.

- **Endogenous effect and contextual effect**

The endogenous effect, where an agent's choice/outcome may depend on those of his/her friends on the same activity, is captured by the coefficients λ_{11} and λ_{22}. The contextual effect, where an agent's choice/outcome may depend on the exogenous characteristics of his/her friends, is captured by γ_1 and γ_2.

The non-identification of a social interaction model caused by the coexistence of those two effects is known as the "reflection problem" (Manski, 1993). For example, in a linear-in-means model, where an agent is equally affected by all the other agents in the network and by nobody outside the network, the mean of endogenous regressor is perfectly collinear with the exogenous regressors. Hence, endogenous and contextual effects cannot be separately identified.

In reality, an agent may not be evenly influenced by all the other agents in a network. In a network model, it is usually assumed that an agent is only influenced by his/her friends. Note that, if individuals i, j are friends and j, k are friends, it does not necessarily imply that i, k are also friends. Thus, the intransitivity in network connections provides an exclusion restriction to identify the model. Bramoullé et al. (2009) show that if intransitivities exist in a network so that I, G^*, G^{*2} are linearly independent, then the characteristics of an agent's second-order (indirect) friends $G^{*2}X$ can be used as instruments to identify the endogenous effect from the contextual effect in the *local-average* model.[5]

On the other hand, when G_r does not have constant row sums, the number of friends represented by $G_r l_{n_r}$ varies across agents. For a *local-aggregate* model, Liu and Lee (2010) show that the Bonacich (1987) centrality, which has $G_r l_{n_r}$ as the leading-order term, can also be used as an instrument for the endogenous effect. For the a *local-aggregate* seemingly unrelated regression (SUR) network model with fixed network effect, we show in the following section that identification is possible through the intransitivity in network connections and/or the variation in Bonacich centrality.

- **Simultaneity effect and cross-activity peer effect**

The simultaneity effect, where an agent's choice/outcome of an activity may depend on his/her choice/outcome of a related activity, can be seen in the coefficients ϕ_1 and ϕ_2. The cross-activity

[5] A stronger identification condition is needed if the network fixed effect is also included in the model.

peer effect, where an agent's choice/outcome may depend on those of his/her friends on a related activity, is represented by the coefficients λ_{21} and λ_{12}.

For a standard simultaneous equations model without social interaction effects, the simultaneity problem is a well known problem for the identification and the usual remedy is to impose exclusion restrictions on the exogenous variables. Cohen-Cole et al. (2012) show that, with the simultaneity effect *or* the cross-activity peer effect (but not both), the *local-average* network model can be identified without imposing any exclusion restrictions on X, as long as J, JG^*, G^{*2}, G^{*3} are linearly independent. In this paper, we show that, by exploiting the variation in Bonacich centrality, the *local-aggregate* network model with the simultaneity effect *or* the cross-activity peer effect can be identified under weaker conditions.

However, either the intransitivity in G or the variation in Bonacich centrality would not be enough to identify the simultaneous equations network model with both simultaneity and cross-activity peer effects. One possible approach to achieve identification is to impose exclusion restrictions on X. We show that, with exclusion restrictions on X, the *local-aggregate* network model with both simultaneity and cross-activity peer effects can be identified under weaker conditions than the *local-average* model.

- **Network correlated effect and cross-activity correlated effect**

Furthermore, the structure of the simultaneous equations network model is flexible enough to allow us to incorporate two types of correlated effects.

First, the network fixed effect given by $\alpha_{1,r}$ and $\alpha_{2,r}$ captures *the network correlated effect* where agents in the same network may behave similarly as they have similar unobserved individual characteristics or they face similar institutional environment. Therefore, the network fixed effect serves as a (partial) remedy for the selection bias that originates from the possible sorting of agents with similar unobserved characteristics into a network.

Second, in the simultaneous equations network model, the error terms of the same agent is allowed to be correlated across equations. The correlation structure of the error term captures *the cross-activity correlated effect* so that the choices/outcomes of the same agent on related activities could be correlated. As our identification results are based on the mean of reduce-form equations, they are not affected by the correlation structure of the error term. However, for estimation efficiency, it is important to take into account the correlation in the disturbances. The estimators proposed in this paper extend the generalized spatial 3SLS estimator in Kelejian and Prucha (2004) to estimate the simultaneous equations network model in the presence of many instruments.

4 Identification Results

Among the regularity assumptions listed in Appendix A, Assumption 4 is a sufficient condition for identification of the simultaneous equations network model. Let Z_1 and Z_2 denote the matrices of right-hand-side (RHS) variables of (14) and (15). For Assumption 4 to hold, $\mathrm{E}(JZ_1)$ and $\mathrm{E}(JZ_2)$ need to have full column rank for large enough n. In this section, we provide sufficient conditions for $\mathrm{E}(JZ_1)$ to have full column rank. The sufficient conditions for $\mathrm{E}(JZ_2)$ to have full column rank can be analogously derived.

In this paper, we focus on the case where G_r does not have constant row sums for some network r. When G_r has constant column sums for all r, the equilibrium implication of the local-aggregate network game is similar to that of the local-average network game (see footnote 4) and the identification conditions are analogous to those given in Cohen-Cole et al. (2012). Henceforth, let ρ and η (possibly with subscripts) denote some generic constant scalars that may take different values for different uses.

4.1 Identification of the SUR network model

First, we consider the seemingly unrelated regression (SUR) network model where $\phi_1 = \phi_2 = \lambda_{21} = \lambda_{12} = 0$. Thus, (4) and (5) become

$$Y_1 = \lambda_{11} G Y_1 + X\beta_1 + GX\gamma_1 + L\alpha_1 + \epsilon_1, \tag{22}$$

$$Y_2 = \lambda_{22} G Y_2 + X\beta_2 + GX\gamma_2 + L\alpha_2 + \epsilon_2. \tag{23}$$

For the SUR network model, an agent's choice/outcome is still allowed to be correlated with his/her own choices/outcomes of related activities through the correlation structure of the disturbances. When $\phi_1 = \phi_2 = \lambda_{21} = \lambda_{12} = 0$, it follows from the reduced-form equation (16) that[6]

$$\mathrm{E}(Y_1) = (I - \lambda_{11} G)^{-1}(X\beta_1 + GX\gamma_1 + L\alpha_1). \tag{24}$$

For (22), let $Z_1 = [GY_1, X, GX]$. Identification of (22) requires $\mathrm{E}(JZ_1) = [\mathrm{E}(JGY_1), JX, JGX]$ to have full column rank. As $(I-\lambda_{11}G)^{-1} = I + \lambda_{11}G(I-\lambda_{11}G)^{-1}$ and $(I-\lambda_{11}G)^{-1}G = G(I-\lambda_{11}G)^{-1}$,

[6] For ease of presentation, we assume G, X are nonstochastic. This assumption can be easily relaxed and the results will be conditional on G, X.

it follows from (24) that

$$\mathrm{E}(JGY_1) = JGX\beta_1 + JG^2(I - \lambda_{11}G)^{-1}X(\lambda_{11}\beta_1 + \gamma_1) + JG(I - \lambda_{11}G)^{-1}L\alpha_1.$$

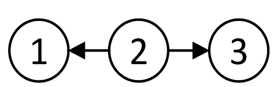
If $\lambda_{11}\beta_1+\gamma_1 \neq 0$, $JG^2(I-\lambda_{11}G)^{-1}X$, with the leading order term JG^2X, can be used as instruments for the endogenous regressor JGY_1. On the other hand, if G_r does not have constant row sums for all $r = 1, \cdots, \bar{r}$, then $JG(I - \lambda_{11}G)^{-1}L \neq 0$. As pointed out by Liu and Lee (2010), the Bonacich centrality given by $G(I - \lambda_{11}G)^{-1}L$, with the leading order term GL, can be used as additional instruments for identification. The following proposition gives a sufficient condition for $\mathrm{E}(JZ_1)$ to have full column rank.

Proposition 1 *Suppose G_r has non-constant row sums for some network r. For equation (22),* $\mathrm{E}(JZ_1)$ *has full column rank if*

(i) $\alpha_{1,r} \neq 0$ *or* $\lambda_{11}\beta_1 + \gamma_1 \neq 0$, *and* I_{n_r}, G_r, G_r^2 *are linearly independent; or*

(ii) $\alpha_{1,r} \neq 0$ *and* $G_r^2 = \rho_1 I_{n_r} + \rho_2 G_r$ *for* $1 - \rho_2\lambda_{11} - \rho_1\lambda_{11}^2 \neq 0$.

The identification condition for the *local-aggregate* SUR model given in Proposition 1 is weaker than that for the *local-average* SUR model. As pointed out by Bramoullé et al. (2009),[7] identification of the local-average model requires the linear independence of I, G^*, G^{*2}, G^{*3}. Consider a data set with $\bar{r}$ networks, where all networks in the data are represented by the graph in Figure 1. For the $\bar{r}$ networks, $G = \mathrm{diag}\{G_r\}_{r=1}^{\bar{r}}$ where G_r is given by the adjacency matrix in Figure 1. For the row normalized adjacency matrix G^*, it is easy to see that $G^{*3} = G^*$. Therefore, it follows by Proposition 5 of Bramoullé et al. (2009) that the local-average SUR model is not identified. On the other hand, for the network in Figure 1, G_r has non-constant row sums and I_4, G_r, G_r^2 are linearly independent. Hence, the local-aggregate SUR model can be identified by Proposition 1(i).



$$G_r = \begin{bmatrix} 0 & 0 & 0 \\ 1 & 0 & 1 \\ 0 & 0 & 0 \end{bmatrix}$$

Figure 2: an example where the local-aggregate model can be identified by Proposition 1(ii).

[7] As the identification conditions given in Bramoullé et al. (2009) are based on the mean of reduce-form equations, they are not affected by the correlation structure of the error term. Hence, they can be applied to the SUR model.

Figure 2 gives an example where the local-aggregate SUR model can be identified by Proposition 1(ii). For the directed graph in Figure 2, the adjacency matrix $G_r = [g_{ij,r}]$, where $g_{ij,r} = 1$ if an arc is directed from i to j and $g_{ij,r} = 0$ otherwise. It is easy to see that $G_r^2 = \rho_1 I_3 + \rho_2 G_r$ for $\rho_1 = \rho_2 = 0$. As the row sums of G_r are not constant, the local-aggregate SUR model is identified.

The following corollary shows that for a network with symmetric adjacency matrix G_r, the local-aggregate SUR model can be identified if G_r has non-constant row sums.

Corollary 1 *Suppose $\lambda_{11}\beta_1 + \gamma_1 \neq 0$ or $\alpha_{1,r} \neq 0$ for some network r. Then, for equation (22),* $\mathrm{E}(JZ_1)$ *has full column rank if G_r is symmetric and has non-constant row sums.*

4.2 Identification of the simultaneous equations network model

4.2.1 Identification under the restrictions $\lambda_{21} = \lambda_{12} = 0$

For the simultaneous equations network model, besides the endogenous, contextual and correlated effects, first we incorporate the simultaneity effect so that an agent's choice/outcome is allowed to depend on his/her own choice/outcome of a related activity. However, we assume that friends' choices/outcomes of related activities have no influence on an agent's choice/outcome. Under the restrictions $\lambda_{21} = \lambda_{12} = 0$, (14) and (15) become

$$Y_1 = \phi_1 Y_2 + \lambda_{11} G Y_1 + X\beta_1 + GX\gamma_1 + L\alpha_1 + \epsilon_1, \quad (25)$$

$$Y_2 = \phi_2 Y_1 + \lambda_{22} G Y_2 + X\beta_2 + GX\gamma_2 + L\alpha_2 + \epsilon_2. \quad (26)$$

For (25), let $Z_1 = [Y_2, GY_1, X, GX]$. For identification of the simultaneous equations model, $\mathrm{E}(JZ_1)$ is required to have full column rank for large enough n. The following proposition gives sufficient conditions for $\mathrm{E}(JZ_1)$ to have full column rank.

Proposition 2 *Suppose G_r has non-constant row sums and I_{n_r}, G_r, G_r^2 are linearly independent for some network r. When $[l_{n_r}, G_r l_{n_r}, G_r^2 l_{n_r}]$ has full column rank,* $\mathrm{E}(JZ_1)$ *of equation (25) has full column rank if*

(i) *$I_{n_r}, G_r, G_r^2, G_r^3$ are linearly independent and D_1 given by (43) has full rank; or*

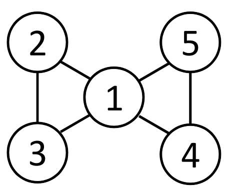
(ii) *$G_r^3 = \rho_1 I_{n_r} + \rho_2 G_r + \rho_3 G_r^2$ and D_1^* given by (44) has full rank.*

When $G_r^2 l_{n_r} = \eta_1 l_{n_r} + \eta_2 G_r l_{n_r}$, $\mathrm{E}(JZ_1)$ *of equation (25) has full column rank if*

(iii) *$I_{n_r}, G_r, G_r^2, G_r^3$ are linearly independent and $D_1^\dagger$ given by (45) has full rank; or*

(iv) *$G_r^3 = \rho_1 I_{n_r} + \rho_2 G_r + \rho_3 G_r^2$ and $D_1^\ddagger$ given by (46) has full rank.*

Cohen-Cole et al. (2012) show that a sufficient condition to identify the *local-average* simultaneous equations model under the restrictions $\lambda_{21} = \lambda_{12} = 0$ requires that $J, JG^*, JG^{*2}, JG^{*3}$ are linearly independent. The sufficient conditions to identify the restricted *local-aggregate* simultaneous equations model given by Proposition 2 is weaker. Consider a data set, where all networks are given by the graph in Figure 3. It is easy to see that, for the row-normalized adjacency matrix $G^* = \text{diag}\{G_r^*\}_{r=1}^{\bar{r}}$, where G_r^* is given in Figure 3, $G^{*3} = -\frac{1}{4}I + \frac{1}{4}G^* + G^{*2}$. Therefore, the condition to identify the local-average model does not hold. On the other hand, for the rth network in the data, $G_r^2 l_5 = 4l_5 + Gl_5$. As the row sums of G_r are not constant and I_5, G_r, G_r^2, G_r^3 are linearly independent, the local-aggregate model can be identified by Proposition 2(iii).

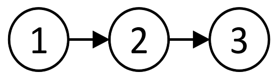


$$G_r = \begin{bmatrix} 0 & 1 & 1 & 1 & 1 \\ 1 & 0 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1 & 0 \end{bmatrix} \qquad G_r^* = \begin{bmatrix} 0 & 1/4 & 1/4 & 1/4 & 1/4 \\ 1/2 & 0 & 1/2 & 0 & 0 \\ 1/2 & 1/2 & 0 & 0 & 0 \\ 1/2 & 0 & 0 & 0 & 1/2 \\ 1/2 & 0 & 0 & 1/2 & 0 \end{bmatrix}$$

Figure 3: an example where the local-aggregate model can be identified by Proposition 2(iii).

For another example where identification is possible for the local-aggregate model but not for the local-average model, let us revisit the network given by the graph in Figure 1. For a data set with $\bar{r}$ such networks, as $G^{*3} = G^*$, the condition to identify the local-average simultaneous equations model given by Cohen-Cole et al. (2012) does not hold. On the other hand, for the adjacency matrix without row-normalization, $G_r^3 = 3G_r$ and $G_r^2 l_4 = 3l_4$. As the row sums of G_r are not constant and I_4, G_r, G_r^2 are linearly independent, the local-aggregate simultaneous equations model can be identified by Proposition 2(iv).

Figure 4 provides an example where the conditions of Proposition 2(ii) are satisfied. For the directed network in Figure 4, $G_r^3 = 0$. As $[l_3, G_r l_3, G_r^2 l_3]$ has full column rank and I_3, G_r, G_r^2 are linearly independent, the local-aggregate simultaneous equations model can be identified by Proposition 2(ii).



$$G_r = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{bmatrix}$$

Figure 4: an example where the local-aggregate model can be identified by Proposition 2(ii).

4.2.2 Identification under the restrictions $\phi_1 = \phi_2 = 0$

Next, let us consider the simultaneous equations model where an agent's choice/outcome is allowed to depend on his/her friends' choices/outcomes of the same activity and a related activity. This specification incorporates the endogenous, contextual, correlated, and cross-activity peer effects, but excludes the simultaneity effect. Under the restrictions $\phi_1 = \phi_2 = 0$, (14) and (15) become

$$Y_1 = \lambda_{11}GY_1 + \lambda_{21}GY_2 + X\beta_1 + GX\gamma_1 + L\alpha_1 + \epsilon_1, \quad (27)$$

$$Y_2 = \lambda_{22}GY_2 + \lambda_{12}GY_1 + X\beta_2 + GX\gamma_2 + L\alpha_2 + \epsilon_2. \quad (28)$$

For (27), let $Z_1 = [GY_1, GY_2, X, GX]$. The following proposition gives sufficient conditions for $\mathrm{E}(JZ_1)$ to have full column rank.

Proposition 3 *Suppose G_r has non-constant row sums and I_{n_r}, G_r, G_r^2 are linearly independent for some network r. When $[l_{n_r}, G_r l_{n_r}, G_r^2 l_{n_r}]$ has full column rank, $\mathrm{E}(JZ_1)$ of equation (27) has full column rank if*

(i) *$I_{n_r}, G_r, G_r^2, G_r^3$ are linearly independent and D_2 given by (47) has full rank; or*

(ii) *$G_r^3 = \rho_1 I_{n_r} + \rho_2 G_r + \rho_3 G_r^2$ and D_2^* given by (48) has full rank.*

When $G_r^2 l_{n_r} = \eta_1 l_{n_r} + \eta_2 G_r l_{n_r}$, $\mathrm{E}(JZ_1)$ of equation (27) has full column rank if

(iii) *$I_{n_r}, G_r, G_r^2, G_r^3$ are linearly independent and $D_2^\dagger$ given by (49) has full rank; or*

(iv) *$G_r^3 = \rho_1 I_{n_r} + \rho_2 G_r + \rho_3 G_r^2$ and $D_2^\ddagger$ given by (50) has full rank.*

For the *local-average* simultaneous equations model under the restrictions $\phi_1 = \phi_2 = 0$, Cohen-Cole et al. (2012) give a sufficient identification condition that requires $J, JG^*, JG^{*2}, JG^{*3}$ to be linearly independent. The sufficient identification condition for the *local-aggregate* model given by Proposition 3 is weaker. As explained in the preceding subsection, for the network given by the graph in Figure 1 or Figure 3, the identification condition for the local-average model does not hold. On the other hand, if G_r is given by Figure 3, the local-aggregate model can be identified by Proposition 3(iii) since the row sums of G_r are not constant, $G_r^2 l_5 = 4l_5 + Gl_5$, and I_5, G_r, G_r^2, G_r^3 are linearly independent. Similarly, if G_r is given by Figure 1, the local-aggregate model can be identified by Proposition 3(iv) since the row sums of G_r are not constant, $G_r^2 l_4 = 3l_4$, $G_r^3 = 3G_r$, and I_4, G_r, G_r^2 are linearly independent.

4.2.3 Non-identification of the general simultaneous equations model

For the general simultaneous equations model given by (14) and (15), the various social interaction effects cannot be separately identified through the mean of the RHS variables without imposing any exclusion restrictions. This is because $\mathrm{E}(\bar{Z}_1)$ and $\mathrm{E}(\bar{Z}_2)$, where $\bar{Z}_1 = [Y_2, GY_1, GY_2, X, GX, L]$ and $\bar{Z}_2 = [Y_2, GY_1, GY_2, X, GX, L]$, do not have full column rank as shown in the following proposition.

Proposition 4 *For (14) and (15),* $\mathrm{E}(\bar{Z}_1)$ *and* $\mathrm{E}(\bar{Z}_2)$ *do not have full column rank.*

Proposition 4 shows that, for the general simultaneous equations model with both simultaneity and cross-activity peer effects, exploiting the intransitivities in social connections and/or variations in Bonacich centrality does not provide enough exclusion restrictions for identification. One way to achieve identification is to impose exclusion restrictions on the coefficients of exogenous variables. Consider the following model

$$Y_1 = \phi_1 Y_2 + \lambda_{11} GY_1 + \lambda_{21} GY_2 + X_1\beta_1 + GX_1\gamma_1 + L\alpha_1 + \epsilon_1, \tag{29}$$

$$Y_2 = \phi_2 Y_1 + \lambda_{22} GY_2 + \lambda_{12} GY_1 + X_2\beta_2 + GX_2\gamma_2 + L\alpha_2 + \epsilon_2, \tag{30}$$

where, for ease of presentation, we assume X_1, X_2 are vectors and $[X_1, X_2]$ has full column rank.[8] From the reduced-form equations (16) and (17), we have

$$\begin{aligned} \mathrm{E}(Y_1) = {} & S^{-1}[X_1\beta_1 + GX_1(\gamma_1 - \lambda_{22}\beta_1) - G^2X_1\lambda_{22}\gamma_1 + X_2\phi_1\beta_2 + GX_2(\lambda_{21}\beta_2 + \phi_1\gamma_2) + G^2X_2\lambda_{21}\gamma_2 \\ & + L(\alpha_1 + \phi_1\alpha_2) + GL(\lambda_{21}\alpha_2 - \lambda_{22}\alpha_1)], \end{aligned} \tag{31}$$

$$\begin{aligned} \mathrm{E}(Y_2) = {} & S^{-1}[X_2\beta_2 + GX_2(\gamma_2 - \lambda_{11}\beta_2) - G^2X_2\lambda_{11}\gamma_2 + X_1\phi_2\beta_1 + GX_1(\lambda_{12}\beta_1 + \phi_2\gamma_1) + G^2X_1\lambda_{12}\gamma_1 \\ & + L(\alpha_2 + \phi_2\alpha_1) + GL(\lambda_{12}\alpha_1 - \lambda_{11}\alpha_2)], \end{aligned} \tag{32}$$

[8] When X_1, X_2 are matrices, we need $[X_1, X_2]$ to have higher column rank than both X_1 and X_2.

where S given by (6). For (29), let $Z_1 = [Y_2, GY_1, GY_2, X_1, GX_1]$. The following proposition gives sufficient conditions for $\mathrm{E}(JZ_1)$ to have full column rank.

Proposition 5 *Suppose G_r has non-constant row sums for some network r. When $[l_{n_r}, G_r l_{n_r}, G_r^2 l_{n_r}]$ has full column rank,* $\mathrm{E}(JZ_1)$ *of equation (29) has full column rank if*

(i) *$I_{n_r}, G_r, G_r^2, G_r^3$ are linearly independent and D_3 given by (51) has full rank; or*

(ii) *I_{n_r}, G_r, G_r^2 are linearly independent, $G_r^3 = \rho_1 I_{n_r} + \rho_2 G_r + \rho_3 G_r^2$ and D_3^* given by (52) has full rank.*

When $G_r^2 l_{n_r} = \eta_1 l_{n_r} + \eta_2 G_r l_{n_r}$, $\mathrm{E}(JZ_1)$ *of equation (29) has full column rank if*

(iii) *$I_{n_r}, G_r, G_r^2, G_r^3$ are linearly independent and $D_3^\dagger$ given by (53) has full rank;*

(iv) *I_{n_r}, G_r, G_r^2 are linearly independent, $G_r^3 = \rho_1 I_{n_r} + \rho_2 G_r + \rho_3 G_r^2$ and $D_3^\ddagger$ given by (54) has full rank; or*

(v) *$G_r^2 = \eta_1 I_{n_r} + \eta_2 G_r$ and $D_3^\sharp$ given by (55) has full rank.*

For the general *local-average* simultaneous equations model, Cohen-Cole et al. (2012) provide a sufficient identification condition that requires J, JG^*, JG^{*2} to be linearly independent. Suppose $G^* = \mathrm{diag}\{G_r^*\}_{r=1}^{\bar{r}}$ where G_r^* is given by Figure 1. It is easy to see that $JG^{*2} = -JG^*$. Therefore, the identification condition for the local-average model does not hold. On the other hand, as G_r given by Figure 1 has non-constant row sums and I_4, G_r, G_r^2 are linearly independent, the local-aggregate model given by (29) and (30) can be identified by Proposition 5(iv).

5 Estimation

5.1 The 2SLS estimator with many instruments

The general simultaneous equations model given by (29) and (30) can be written more compactly as

$$Y_1 = Z_1\delta_1 + L\alpha_1 + \epsilon_1 \quad \text{and} \quad Y_2 = Z_2\delta_2 + L\alpha_2 + \epsilon_2, \tag{33}$$

where $Z_1 = [Y_2, GY_1, GY_2, X_1, GX_1]$, $Z_2 = [Y_1, GY_2, GY_1, X_2, GX_2]$, $\delta_1 = (\phi_1, \lambda_{11}, \lambda_{21}, \beta_1', \gamma_1')'$, and $\delta_2 = (\phi_2, \lambda_{22}, \lambda_{12}, \beta_2', \gamma_2')'$. As $JL = 0$, the within transformation with projector J gives $JY_1 = JZ_1\delta_1 + J\epsilon_1$ and $JY_2 = JZ_2\delta_2 + J\epsilon_2$. From the reduced-form equations (16) and (17), we have

$$JZ_1 = \mathrm{E}(JZ_1) + U_1 = F_1 + U_1, \quad \text{and} \quad JZ_2 = \mathrm{E}(JZ_2) + U_2 = F_2 + U_2, \tag{34}$$

where $F_1 = J[\mathrm{E}(Y_2), G\mathrm{E}(Y_1), G\mathrm{E}(Y_2), X_1, GX_1]$, $F_2 = J[\mathrm{E}(Y_1), G\mathrm{E}(Y_2), G\mathrm{E}(Y_1), X_2, GX_2]$, $U_1 = J[S^{-1}u_2, GS^{-1}u_1, GS^{-1}u_2, 0]$, and $U_2 = J[S^{-1}u_1, GS^{-1}u_2, GS^{-1}u_1, 0]$, with $\mathrm{E}(Y_1), \mathrm{E}(Y_2)$ given by (31) and (32), S given by (6), and u_1, u_2 given by (18) and (19).

Based on (34), the best instruments for JZ_1 and JZ_2 are F_1 and F_2 respectively (Lee, 2003). However, both F_1 and F_2 are infeasible as they involve unknown parameters. Hence, we use linear combinations of feasible instruments to approximate F_1 and F_2 as in Kelejian and Prucha (2004) and Liu and Lee (2010). Let $\bar{G} = \phi_1\phi_2 I + (\lambda_{11}+\lambda_{22}+\phi_1\lambda_{12}+\phi_2\lambda_{21})G - (\lambda_{11}\lambda_{22}-\lambda_{12}\lambda_{21})G^2$. Under some regularity conditions (see footnote 2), we have $||\bar{G}||_\infty < 1$. Then, $S^{-1} = (I-\bar{G})^{-1} = \sum_{j=0}^{\infty} \bar{G}^j = \sum_{j=0}^{p} \bar{G}^j + \bar{G}^{p+1}S^{-1}$. It follows that $||S^{-1} - \sum_{j=0}^{p} \bar{G}^j||_\infty \leq ||\bar{G}||_\infty^{p+1}||S^{-1}||_\infty$. As $||\bar{G}||_\infty < 1$, the approximation error of $\sum_{j=0}^{p} \bar{G}^j$ diminishes in a geometric rate as $p \to \infty$. Since $\sum_{j=0}^{p} \bar{G}^j$ can be considered as a linear combination of $[I, G, \cdots, G^{2p}]$, the best instruments F_1 and F_2 can be approximated by a linear combination of $n \times K$ IV matrix

$$Q_K = J[X_1, GX_1, \cdots, G^{2p+3}X_1, X_2, GX_2, \cdots, G^{2p+3}X_2, GL, \cdots, G^{2p+2}L] \tag{35}$$

with an approximation error diminishing very fast when K (or p) goes to infinity, as required by Assumption 5 in Appendix A. Let $P_K = Q_K(Q_K'Q_K)^{-}Q_K'$. The many-instrument two-stage least-squares (2SLS) estimators for δ_1 and δ_2 are $\hat{\delta}_{1,2sls} = (Z_1'P_KZ_1)^{-1}Z_1'P_KY_1$ and $\hat{\delta}_{2,2sls} = (Z_2'P_KZ_2)^{-1}Z_2'P_KY_2$.

Let $H_{11} = \lim_{n\to\infty} \frac{1}{n}F_1'F_1$ and $H_{22} = \lim_{n\to\infty} \frac{1}{n}F_2'F_2$. The following proposition establishes the consistency and asymptotic normality of the many-instrument 2SLS estimator.

Proposition 6 *Under Assumptions 1-5, if $K \to \infty$ and $K/n \to 0$, then $\sqrt{n}(\hat{\delta}_{1,2sls} - \delta_1 - b_{1,2sls}) \xrightarrow{d} N(0, \sigma_1^2 H_{11}^{-1})$ and $\sqrt{n}(\hat{\delta}_{2,2sls} - \delta_2 - b_{2,2sls}) \xrightarrow{d} N(0, \sigma_2^2 H_{22}^{-1})$, where $b_{1,2sls} = (Z_1'P_KZ_1)^{-1}\mathrm{E}(U_1'P_K\epsilon_1) = O_p(K/n)$ and $b_{2,2sls} = (Z_2'P_KZ_2)^{-1}\mathrm{E}(U_2'P_K\epsilon_2) = O_p(K/n)$.*

From Proposition 6, when the number of instruments K grows at a rate slower than the sample size n, the 2SLS estimators are consistent and asymptotically normal. However, the asymptotic distribution of the 2SLS estimator may not center around the true parameter value due to the presence of many-instrument bias of order $O_p(K/n)$ (see, e.g., Bekker, 1994). If $K^2/n \to 0$, then $\sqrt{n}b_{1,2sls} \xrightarrow{p} 0$, $\sqrt{n}b_{2,2sls} \xrightarrow{p} 0$ and the 2SLS estimators are properly centered.

The condition that $K/n \to 0$ is crucial for the 2SLS estimator to be consistent. To see this, we look at the first-order conditions of the 2SLS, $\frac{1}{n}Z_1'P_K(Y_1 - Z_1\hat{\delta}_{1,2sls}) = 0$ and $\frac{1}{n}Z_2'P_K(Y_2 -$

$Z_2\hat{\delta}_{2,2sls}) = 0$. At the true parameter values, $\mathrm{E}[\frac{1}{n}Z_1'P_K(Y_1 - Z_1\delta_1)] = \frac{1}{n}\mathrm{E}(U_1'P_K\epsilon_1)$ and $\mathrm{E}[\frac{1}{n}Z_2'P_K(Y_2 - Z_2\delta_2)] = \frac{1}{n}\mathrm{E}(U_2'P_K\epsilon_2)$, where

$$\mathrm{E}(U_1'P_K\epsilon_1) = \begin{bmatrix} (\sigma_{12} + \phi_2\sigma_1^2)\mathrm{tr}(P_K S^{-1}) + (\lambda_{12}\sigma_1^2 - \lambda_{11}\sigma_{12})\mathrm{tr}(P_K S^{-1}G) \\ (\sigma_1^2 + \phi_1\sigma_{12})\mathrm{tr}(P_K GS^{-1}) + (\lambda_{21}\sigma_{12} - \lambda_{22}\sigma_1^2)\mathrm{tr}(P_K GS^{-1}G) \\ (\sigma_{12} + \phi_2\sigma_1^2)\mathrm{tr}(P_K GS^{-1}) + (\lambda_{12}\sigma_1^2 - \lambda_{11}\sigma_{12})\mathrm{tr}(P_K GS^{-1}G) \\ 0_{2k_x\times 1} \end{bmatrix} = O(K) \tag{36}$$

$$\mathrm{E}(U_2'P_K\epsilon_2) = \begin{bmatrix} (\sigma_{12} + \phi_1\sigma_2^2)\mathrm{tr}(P_K S^{-1}) + (\lambda_{21}\sigma_2^2 - \lambda_{22}\sigma_{12})\mathrm{tr}(P_K S^{-1}G) \\ (\sigma_2^2 + \phi_2\sigma_{12})\mathrm{tr}(P_K GS^{-1}) + (\lambda_{12}\sigma_{12} - \lambda_{11}\sigma_2^2)\mathrm{tr}(P_K GS^{-1}G) \\ (\sigma_{12} + \phi_1\sigma_2^2)\mathrm{tr}(P_K GS^{-1}) + (\lambda_{21}\sigma_2^2 - \lambda_{22}\sigma_{12})\mathrm{tr}(P_K GS^{-1}G) \\ 0_{2k_x\times 1} \end{bmatrix} = O(K) \tag{37}$$

by Lemma C.2 in the Appendix. Therefore, $\mathrm{E}[\frac{1}{n}Z_1'P_K(Y_1 - Z_1\delta_1)]$ and $\mathrm{E}[\frac{1}{n}Z_2'P_K(Y_2 - Z_2\delta_2)]$ may not converge to zero and, thus, the 2SLS estimators may not be consistent, if the number of instruments grows at the same or a faster rate than the sample size.

Note that the submatrix GL in the IV matrix Q_K given by (35) has $\bar{r}$ columns, where $\bar{r}$ is the number of networks. Hence, $K/n \to 0$ implies $\bar{r}/n = 1/\bar{m} \to 0$, where $\bar{m}$ is the average network size. So for the 2SLS estimator with the IV matrix Q_K to be consistent, the average network size needs to be large. On the other hand, $K^2/n \to 0$ implies $\bar{r}^2/n = \bar{r}/\bar{m} \to 0$. So for the 2SLS estimator to be properly centered, the average network size needs to be large relative to the number of networks.

The many-instrument bias of the 2SLS estimator can be corrected by the estimated leading-order biases $b_{1,2sls}$ and $b_{2,2sls}$ given in Proposition 6. Let $\tilde{\delta}_1 = (\tilde{\phi}_1, \tilde{\lambda}_{11}, \tilde{\lambda}_{21}, \tilde{\beta}_1', \tilde{\gamma}_1')'$ and $\tilde{\delta}_2 = (\tilde{\phi}_2, \tilde{\lambda}_{22}, \tilde{\lambda}_{12}, \tilde{\beta}_2', \tilde{\gamma}_2')'$ be $\sqrt{n}$-consistent preliminary 2SLS estimators based on a fixed number of instruments (e.g., $Q = J[X_1, GX_1, G^2X_1, X_2, GX_2, G^2X_2]$). Let $\tilde{\epsilon}_1 = J(Y_1 - Z_1\tilde{\delta}_1)$ and $\tilde{\epsilon}_2 = J(Y_2 - Z_2\tilde{\delta}_2)$. The leading-order biases can be estimated by $\hat{b}_{1,2sls} = (Z_1'P_K Z_1)^{-1}\hat{\mathrm{E}}(U_1'P_K\epsilon_1)$ and $\hat{b}_{2,2sls} = (Z_2'P_K Z_2)^{-1}\hat{\mathrm{E}}(U_2'P_K\epsilon_2)$, where $\hat{\mathrm{E}}(U_1'P_K\epsilon_1)$ and $\hat{\mathrm{E}}(U_2'P_K\epsilon_2)$ are obtained by replacing the unknown parameters in (36) and (37) by $\tilde{\delta}_1, \tilde{\delta}_2$ and

$$\tilde{\sigma}_1^2 = \tilde{\epsilon}_1'\tilde{\epsilon}_1/(n - \bar{r}), \quad \tilde{\sigma}_1^2 = \tilde{\epsilon}_2'\tilde{\epsilon}_2/(n - \bar{r}), \quad \hat{\sigma}_{12} = \tilde{\epsilon}_1'\tilde{\epsilon}_2/(n - \bar{r}). \tag{38}$$

The bias-corrected 2SLS (BC2SLS) estimators are given by $\hat{\delta}_{1,bc2sls} = \hat{\delta}_{1,2sls} - \hat{b}_{1,2sls}$ and $\hat{\delta}_{2,bc2sls} =$

$\hat{\delta}_{2,2sls} - \hat{b}_{2,2sls}$.

Proposition 7 *Under Assumptions 1-5, if $K \to \infty$ and $K/n \to 0$, then* $\sqrt{n}(\hat{\delta}_{1,bc2sls} - \delta_1) \xrightarrow{d} N(0, \sigma_1^2 H_{11}^{-1})$ *and* $\sqrt{n}(\hat{\delta}_{2,bc2sls} - \delta_2) \xrightarrow{d} N(0, \sigma_2^2 H_{22}^{-1})$.

As $K/n \to 0$ implies $1/\bar{m} \to 0$ for the IV matrix Q_K, it follows that the BC2SLS estimators have properly centered asymptotic normal distributions as long as the average network size $\bar{m}$ is large.

5.2 The 3SLS estimator with many instruments

The 2SLS and BC2SLS estimators consider equation-by-equation estimation and are inefficient as they do not make use of the cross-equation correlation in the disturbances. To fully utilize the information in the system, we consider the three-stage least-squares (3SLS) estimator proposed by Kelejian and Prucha (2004) in the presence of many instruments.

We stack the equations in the system (33) as

$$Y = Z\delta + (I_2 \otimes L)\alpha + \epsilon,$$

where $Y = (Y_1', Y_2')'$, $Z = \text{diag}\{Z_1, Z_2\}$, $\delta = (\delta_1', \delta_2')'$, $\alpha = (\alpha_1', \alpha_2')'$, and $\epsilon = (\epsilon_1', \epsilon_2')'$. As $(I_2 \otimes J)(I_2 \otimes L) = 0$, the within transformation with projector J gives $(I_2 \otimes J)Y = (I_2 \otimes J)Z\delta + (I_2 \otimes J)\epsilon$. Let

$$\Sigma = \begin{bmatrix} \sigma_1^2 & \sigma_{12} \\ \sigma_{12} & \sigma_2^2 \end{bmatrix} \quad \text{and} \quad \tilde{\Sigma} = \begin{bmatrix} \tilde{\sigma}_1^2 & \tilde{\sigma}_{12} \\ \tilde{\sigma}_{12} & \tilde{\sigma}_2^2 \end{bmatrix}, \tag{39}$$

where $\tilde{\sigma}_1^2, \tilde{\sigma}_1^2, \tilde{\sigma}_{12}$ are estimated by (38). As $\mathrm{E}(\epsilon\epsilon') = \Sigma \otimes I$, the 3SLS estimator with the IV matrix Q_K is given by $\hat{\delta}_{3sls} = [Z'(\tilde{\Sigma}^{-1} \otimes P_K)Z]^{-1} Z'(\tilde{\Sigma}^{-1} \otimes P_K)Y$.

Let $F = \text{diag}\{F_1, F_2\}$, $U = \text{diag}\{U_1, U_2\}$, and $H = \lim_{n\to\infty} \frac{1}{n} F'(\Sigma^{-1} \otimes I)F$. The following proposition gives the asymptotic distribution of the many-instrument 3SLS estimator.

Proposition 8 *Under Assumptions 1-5, if $K \to \infty$ and $K/n \to 0$, then*

$$\sqrt{n}(\hat{\delta}_{3sls} - \delta - b_{3sls}) \xrightarrow{d} N(0, H^{-1})$$

where $b_{3sls} = [Z'(\Sigma^{-1} \otimes P_K)Z]^{-1}\mathrm{E}[U'(\Sigma^{-1} \otimes P_K)\epsilon] = O_p(K/n)$ *and*

$$\mathrm{E}[U'(\Sigma^{-1} \otimes P_K)\epsilon] = \begin{bmatrix} \phi_2 \mathrm{tr}(P_K S^{-1}) + \lambda_{12}\mathrm{tr}(P_K S^{-1} G) \\ \mathrm{tr}(P_K G S^{-1}) - \lambda_{22}\mathrm{tr}(P_K G S^{-1} G) \\ \phi_2 \mathrm{tr}(P_K G S^{-1}) + \lambda_{12}\mathrm{tr}(P_K G S^{-1} G) \\ 0_{2k_x \times 1} \\ \phi_1 \mathrm{tr}(P_K S^{-1}) + \lambda_{21}\mathrm{tr}(P_K S^{-1} G) \\ \mathrm{tr}(P_K G S^{-1}) - \lambda_{11}\mathrm{tr}(P_K G S^{-1} G) \\ \phi_1 \mathrm{tr}(P_K G S^{-1}) + \lambda_{21}\mathrm{tr}(P_K G S^{-1} G) \\ 0_{2k_x \times 1} \end{bmatrix}. \tag{40}$$

Similar to the 2SLS estimator, when the number of instruments goes to infinity at a rate slower than the sample size, the 3SLS estimator is consistent and asymptotically normal with an asymptotic bias of order $O_p(K/n)$. If $K^2/n \to 0$, then $\sqrt{n} b_{3sls} \xrightarrow{p} 0$ and the 3SLS estimator is properly centered and efficient as the covariance matrix H^{-1} attains the efficiency lower bound for the class of IV estimators.

The leading-order asymptotic bias of the 3SLS estimator given in Proposition 8 can be estimated to correct the many-instrument bias. Let the estimated bias be

$$\hat{b}_{3sls} = [Z'(\tilde{\Sigma}^{-1} \otimes P_K)Z]^{-1}\widehat{\mathrm{E}}[U'(\Sigma^{-1} \otimes P_K)\epsilon],$$

where $\tilde{\Sigma}$ is given by (39) and $\widehat{\mathrm{E}}[U'(\Sigma^{-1} \otimes P_K)\epsilon]$ is obtained by replacing the unknown parameters in (40) by $\sqrt{n}$-consistent preliminary 2SLS estimators $\tilde{\delta}_1$ and $\tilde{\delta}_2$. The bias-corrected 3SLS (BC3SLS) estimator is given by $\hat{\delta}_{bc3sls} = \hat{\delta}_{3sls} - \hat{b}_{3sls}$. The following proposition shows that the BC3SLS estimator is properly centered and asymptotically efficient if the number of instruments increases slower than the sample size.

Proposition 9 *Under Assumptions 1-5, if* $K \to \infty$ *and* $K/n \to 0$*, then*

$$\sqrt{n}(\hat{\delta}_{bc3sls} - \delta) \xrightarrow{d} N(0, H^{-1}).$$

6 Monte Carlo Experiments

To investigate the finite sample performance of the 2SLS and 3SLS estimators, we conduct a limited simulation study based on the following model

$$Y_1 = \phi_1 Y_2 + \lambda_{11} G Y_1 + \lambda_{21} G Y_2 + X_1 \beta_1 + G X_1 \gamma_1 + L\alpha_1 + \epsilon_1, \quad (41)$$

$$Y_2 = \phi_2 Y_1 + \lambda_{22} G Y_2 + \lambda_{12} G Y_1 + X_2 \beta_2 + G X_2 \gamma_2 + L\alpha_2 + \epsilon_2. \quad (42)$$

For the experiment, we consider three samples with different numbers of networks $\bar{r}$ and network sizes m_r. The first sample contains 30 networks with equal sizes of $m_r = 10$. To study the effect of a larger network size, the second sample contains 30 networks with equal sizes of $m_r = 15$. To study the effect of more networks, the third sample contains 60 networks with equal sizes of $m_r = 15$. For each network, the adjacency matrix G_r is generated as follows. First, for the ith row of G_r ($i = 1, \cdots, m_r$), we generate an integer p_{ri} uniformly at random from the set of integers $\{1, 2, 3\}$. Then, if $i + p_{ri} \leq m_r$, we set the $(i+1)$th, $\cdots$, $(i + p_{ri})$th entries of the ith row of G_r to be ones and the other entries in that row to be zeros; otherwise, the entries of ones will be wrapped around such that the first $(p_{ri} - m_r)$ entries of the ith row will be ones.[9]

We conduct 500 repetitions for each specification in this Monte Carlo experiment. In each repetition, for $j = 1, 2$, the $n \times 1$ vector of exogenous variables X_j is generated from $N(0, I)$, and the $\bar{r} \times 1$ vector of network fixed effect coefficients α_j is generated from $N(0, I_{\bar{r}})$. The error terms $\epsilon = (\epsilon_1', \epsilon_2')'$ is generated from $N(0, \Sigma \otimes I)$, where Σ is given by (39). In the data generating process (DGP), we set $\sigma_1^2 = \sigma_2^2 = 1$ and let σ_{12} vary in the experiment. For the other parameters in the model, we set $\phi_1 = \phi_2 = 0.2$, $\lambda_{11} = \lambda_{22} = 0.1$, and $\lambda_{12} = \lambda_{21} = 0.2$.[10] We let β's and γ's vary in the experiment.

We consider the following estimators in the simulation experiment: (i) 2SLS-1 and 3SLS-1 with the IV matrix $Q_1 = J[X_1, GX_1, G^2X_1, X_2, GX_2, G^2X_2]$; (ii) 2SLS-2 and 3SLS-2 with the IV matrix $Q_2 = [Q_1, JGL]$; and (iii) BC2SLS and BC3SLS. The IV matrix Q_1 is based on the exogenous attributes of direct and indirect friends. The IV matrix Q_2 also uses the number of friends given by GL as additional instruments to improve estimation efficiency. Note that GL has $\bar{r}$ columns. So the

[9] Note that the parameter space of ϕ's and λ's depends on $||G||_\infty$ (see footnote 2). If $||G||_\infty$ varies in the experiment, so does the parameter space. To facilitate comparison, we keep $||G||_\infty = 3$ in the experiments. We have tried different values for $||G||_\infty$. The simulation results are similar to those reported here.

[10] We choose ϕ's and λ's so that the S is invertible according to footnote 2.

number of instruments in Q_2 increases with the number of networks.

The estimation results of equation (41) are reported in Tables 1-6. We report the mean and standard deviation (SD) of the empirical distributions of the estimates. To facilitate the comparison of different estimators, we also report their root mean square errors (RMSE). The main observations from the experiment are summarized as follows.

[Tables 1-6 approximately here]

(a) The additional instruments based on the number of friends in Q_2 reduce SDs of 2SLS and 3SLS estimators. When the IVs in Q_1 are strong (i.e., $\beta_1 = \beta_2 = \gamma_1 = \gamma_2 = 0.8$ as in Tables 1-3) and the correlation across equations is weak ($\sigma_{12} = 0.1$), for the sample with $m_r = 10$ and $\bar{r} = 30$ in Table 1, SD reductions of 2SLS-2 estimators of $\phi_1, \lambda_{11}, \lambda_{21}, \beta_1, \gamma_1$ (relative to 2SLS-1) are, respectively, about 5.9%, 14.3%, 9.7%, 3.0%, and 5.5%. As the correlation across equations increases, the SD reduction also increases. When $\sigma_{12} = 0.9$ (see the bottom panel of Table 1), SD reductions of 2SLS-2 estimators of $\phi_1, \lambda_{11}, \lambda_{21}, \beta_1, \gamma_1$ are, respectively, about 15.7%, 20.6%, 16.1%, 9.2%, and 11.1%. Furthermore, the SD reduction is more significant when the IVs in Q_1 are less informative (i.e., $\beta_1 = \beta_2 = \gamma_1 = \gamma_2 = 0.4$ as in Tables 4-6). When $\sigma_{12} = 0.1$ (see the top panel of Table 4), SD reductions of 2SLS-2 estimators of $\phi_1, \lambda_{11}, \lambda_{21}, \beta_1, \gamma_1$ are, respectively, about 20.8%, 43.2%, 36.4%, 7.5%, and 13.6%. The SD reduction of the 3SLS estimator with Q_2 follows a similar pattern.

(b) The additional instruments in Q_2 introduce biases into 2SLS and 3SLS estimators. The size of the bias increases as the correlation across equations σ_{12} increases and as the IVs in Q_1 becomes less informative (i.e., $\beta_1, \beta_2, \gamma_1, \gamma_2$ become smaller). The size of the bias reduces as the network size increases. The impact of the number of networks on the bias is less obvious.

(c) The proposed bias-correction procedure substantially reduces the many-instrument bias for both the 2SLS and 3SLS estimators. For example, in Table 4, bias reductions of BC3SLS estimators of $\phi_1, \lambda_{11}, \lambda_{21}$ are, respectively, 100.0%, 86.7%, and 66.7%, when $\sigma_{12} = 0.1$.

(d) The 3SLS estimator improves the efficiency upon the 2SLS estimator. The improvement is most prominent when the correlation across equations is strong. In Table 1, when $\sigma_{12} = 0.9$, SD reductions of BC3SLS estimators of $\phi_1, \lambda_{11}, \lambda_{21}, \beta_1, \gamma_1$ (relative to BC2SLS) are, respectively, about 8.8%, 6.5%, 3.3%, 21.2%, and 11.1%.

7 Conclusion

In this paper, we consider specification, identification and estimation of network models in a system of simultaneous equations. We show that, with or without row-normalization of the network adjacency matrix, the network model has different equilibrium implications, needs different identification conditions, and requires different estimation strategies. When the network adjacency matrix is not row-normalized, the Bonacich (1987) centrality based on the number of direct and indirect friends of agents in a network can be used to identify social interaction effects and improve estimation efficiency. We derive the identification conditions for some specifications of the simultaneous equations network model with a non-row-normalized adjacency matrix, and show that the identification conditions are weaker than those for the model with a row-normalized adjacency matrix derived by Cohen-Cole et al. (2012).

For efficient estimation, we propose 2SLS and 3SLS estimators for the simultaneous equations network model using a set of feasible instruments to approximate the best (infeasible) instruments given by the reduced-form equations of the model. When the network adjacency matrix is not row-normalized, the set of feasible instruments includes the leading order terms of the Bonacich centrality for each network, and thus the number of instruments depends on the number of networks. When there are many networks in the data, we would have many instruments. We show that the proposed 2SLS and 3SLS estimators are consistent and asymptotically normally distributed (with an asymptotic bias) when the number of instruments increases at a rate slower than the sample size. We also propose a bias-correction procedure based on the estimated leading-order term of the many-instrument bias. The bias-corrected 2SLS and 3SLS estimators have an properly centered asymptotic normal distribution when the number of instruments grows slower than the sample size (or, when the average network size is large). Monte Carlo experiments show that the instruments based on the Bonacich centrality reduce the standard errors of the 2SLS and 3SLS estimators and the bias-corrected estimators perform well with a moderate network size (say, $\bar{m}_r = 10$).

APPENDIX

A Assumptions

In this appendix, we list regularity conditions for the asymptotic properties of the proposed estimators. Henceforth, uniformly bounded in row (column) sums in absolute value of a sequence of square matrices $\{A\}$ will be abbreviated as UBR (UBC), and uniformly bounded in both row and column sums in absolute value as UB.[11]

Assumption 1 The vector of disturbances is given by $\epsilon = (\epsilon_1', \epsilon_2')' = (\Sigma_*' \otimes I_n)v$, where Σ_*' is a nonsingular matrix such that $\Sigma_*'\Sigma_* = \Sigma$ and the elements of v are i.i.d. with zero mean, unit variance and finite fourth moments. Furthermore, the diagonal elements of Σ are bounded by some finite constant.

Assumption 2 The matrix of exogenous (nonstochastic) regressors X has full column rank (for n sufficient large). The elements of X are uniformly bounded in absolute value.

Assumption 3 The matrix S is nonsingular. The sequences of matrices $\{G\}$ and $\{S^{-1}\}$ are UB.

Assumption 4 Let $F_i = \mathrm{E}(JZ_i)$ for $i = 1, 2$, and $F = \mathrm{diag}\{F_1, F_2\}$. Then, $H_{ij} = \lim_{n\to\infty} \frac{1}{n} F_i' F_j$, for $i, j = 1, 2$, and $H = \lim_{n\to\infty} \frac{1}{n} F'(\Sigma^{-1} \otimes I)F$ are finite nonsingular matrices.

Assumption 5 There exist matrices π_1 and π_2 such that, for $i = 1, 2$, $||F_i - Q_K\pi_i||_\infty \to 0$ as $n, K \to \infty$.

Assumption 1-3 originate in Kelejian and Prucha (2004). The matrix of exogenous regressors X is assumed to be nonstochastic for ease of presentation. If X is allowed to be stochastic, then appropriate moment conditions need to be imposed, and the results presented in this paper can be considered as conditional on X instead. Assumption 4 is for the identification of the network model. It also implies the concentration parameter grows at the same rate as the sample size (Liu and Lee, 2010). Assumption 5 requires the (infeasible) best IV matrix F_i (for $i = 1, 2$) can be well approximated by a certain linear combination of the feasible IV matrix Q_K as the number of instruments increases with the sample size. This condition is commonly assumed in the many-instruments literature (see, eg., Donald and Newey, 2001; Hansen et al., 2008; Hausman et al., 2008).

[11] A sequence of square matrices $\{A\}$, where $A = [A_{ij}]$, is said to be UBR (UBC) if the sequence of row sum matrix norm $||A||_\infty = \max_{i=1,\cdots,n} \sum_{j=1}^n |A_{ij}|$ (column sum matrix norm $||A||_1 = \max_{j=1,\cdots,n} \sum_{i=1}^n |A_{ij}|$) is bounded. (Horn and Johnson, 1985)

B Rank Conditions

In this appendix, we list the matrices whose rank conditions are used for the identification of the simultaneous equations model.

Let

$$A_1 = \begin{bmatrix} a_{1,1} \\ a_{2,1} \\ a_{3,1} \\ a_{4,1} \end{bmatrix} = \begin{bmatrix} \phi_2\beta_1 + \beta_2 & 0 & 1-\phi_1\phi_2 & 0 & 0 \\ \phi_2\gamma_1 + \gamma_2 - \lambda_{11}\beta_2 & \beta_1 + \phi_1\beta_2 & -(\lambda_{11}+\lambda_{22}) & 1-\phi_1\phi_2 & 0 \\ -\lambda_{11}\gamma_2 & \gamma_1 + \phi_1\gamma_2 - \lambda_{22}\beta_1 & \lambda_{11}\lambda_{22} & -(\lambda_{11}+\lambda_{22}) & 0 \\ 0 & -\lambda_{22}\gamma_1 & 0 & \lambda_{11}\lambda_{22} & 0 \end{bmatrix},$$

$$C_1 = \begin{bmatrix} c_{1,1} \\ c_{2,1} \\ c_{3,1} \end{bmatrix} = \begin{bmatrix} \phi_2\alpha_1 + \alpha_2 & 0 & 0 & 0 & 1-\phi_1\phi_2 \\ -\lambda_{11}\alpha_2 & \alpha_1 + \phi_1\alpha_2 & 0 & 0 & -(\lambda_{11}+\lambda_{22}) \\ 0 & -\lambda_{22}\alpha_1 & 0 & 0 & \lambda_{11}\lambda_{22} \end{bmatrix}.$$

Let $A_1^* = [(a_{1,1}+\rho_1 a_{4,1})', (a_{2,1}+\rho_2 a_{4,1})', (a_{3,1}+\rho_3 a_{4,1})']'$ and $C_1^* = [(c_{1,1}+\eta_1 c_{3,1})', (c_{2,1}+\eta_2 c_{3,1})']'$. Then,

$$D_1 = [A_1', C_1']' \tag{43}$$

$$D_1^* = [A_1^{*\prime}, C_1']' \tag{44}$$

$$D_1^\dagger = [A_1', C_1^{*\prime}]' \tag{45}$$

$$D_1^\ddagger = [A_1^{*\prime}, C_1^{*\prime}]'. \tag{46}$$

Let

$$A_2 = \begin{bmatrix} a_{1,2} \\ a_{2,2} \\ a_{3,2} \\ a_{4,2} \end{bmatrix} = \begin{bmatrix} 0 & 0 & 1 & 0 & 0 \\ \beta_1 & \beta_2 & -(\lambda_{11}+\lambda_{22}) & 1 & 0 \\ \lambda_{21}\beta_2 - \lambda_{22}\beta_1 + \gamma_1 & \lambda_{12}\beta_1 - \lambda_{11}\beta_2 + \gamma_2 & \lambda_{11}\lambda_{22} - \lambda_{12}\lambda_{21} & -(\lambda_{11}+\lambda_{22}) & 0 \\ \lambda_{21}\gamma_2 - \lambda_{22}\gamma_1 & \lambda_{12}\gamma_1 - \lambda_{11}\gamma_2 & 0 & \lambda_{11}\lambda_{22} - \lambda_{12}\lambda_{21} & 0 \end{bmatrix},$$

$$C_2 = \begin{bmatrix} c_{1,2} \\ c_{2,2} \\ c_{3,2} \end{bmatrix} = \begin{bmatrix} 0 & 0 & 0 & 0 & 1 \\ \alpha_1 & \alpha_2 & 0 & 0 & -(\lambda_{11}+\lambda_{22}) \\ \lambda_{21}\alpha_2 - \lambda_{22}\alpha_1 & \lambda_{12}\alpha_1 - \lambda_{11}\alpha_2 & 0 & 0 & \lambda_{11}\lambda_{22} - \lambda_{12}\lambda_{21} \end{bmatrix}.$$

Let $A_2^* = [(a_{1,2}+\rho_1 a_{4,2})', (a_{2,2}+\rho_2 a_{4,2})', (a_{3,2}+\rho_3 a_{4,2})']'$ and $C_2^* = [(c_{1,2}+\eta_1 c_{3,2})', (c_{2,2}+\eta_2 c_{3,2})']'$. Then,

$$D_2 = [A_2', C_2']' \tag{47}$$

$$D_2^* = [A_2^{*\prime}, C_2']' \tag{48}$$

$$D_2^\dagger = [A_2', C_2^{*\prime}]' \tag{49}$$

$$D_2^\ddagger = [A_2^{*\prime}, C_2^{*\prime}]'. \tag{50}$$

Let

$$A_3 = \begin{bmatrix} a_{1,3} \\ a_{2,3} \\ a_{3,3} \\ a_{4,3} \end{bmatrix} = \begin{bmatrix} \phi_2\beta_1 & 0 & 0 & 1-\phi_1\phi_2 & 0 & 0 \\ \lambda_{12}\beta_1+\phi_2\gamma_1 & \beta_1 & \phi_2\beta_1 & \omega & 1-\phi_1\phi_2 & 0 \\ \lambda_{12}\gamma_1 & \gamma_1-\lambda_{22}\beta_1 & \lambda_{12}\beta_1+\phi_2\gamma_1 & \lambda_{11}\lambda_{22}-\lambda_{12}\lambda_{21} & \omega & 0 \\ 0 & -\lambda_{22}\gamma_1 & \lambda_{12}\gamma_1 & 0 & \lambda_{11}\lambda_{22}-\lambda_{12}\lambda_{21} & 0 \end{bmatrix},$$

$$B_3 = \begin{bmatrix} b_{1,3} \\ b_{2,3} \\ b_{3,3} \\ b_{4,3} \end{bmatrix} = \begin{bmatrix} \beta_2 & 0 & 0 & 0 & 0 & 0 \\ \gamma_2-\lambda_{11}\beta_2 & \phi_1\beta_2 & \beta_2 & 0 & 0 & 0 \\ -\lambda_{11}\gamma_2 & \lambda_{21}\beta_2+\phi_1\gamma_2 & \gamma_2-\lambda_{11}\beta_2 & 0 & 0 & 0 \\ 0 & \lambda_{21}\gamma_2 & -\lambda_{11}\gamma_2 & 0 & 0 & 0 \end{bmatrix},$$

$$C_3 = \begin{bmatrix} c_{1,3} \\ c_{2,3} \\ c_{3,3} \end{bmatrix} = \begin{bmatrix} \alpha_{2,r}+\phi_2\alpha_{1,r} & 0 & 0 & 0 & 0 & 1-\phi_1\phi_2 \\ \lambda_{12}\alpha_{1,r}-\lambda_{11}\alpha_{2,r} & \alpha_{1,r}+\phi_1\alpha_{2,r} & \alpha_{2,r}+\phi_2\alpha_{1,r} & 0 & 0 & \omega \\ 0 & \lambda_{21}\alpha_{2,r}-\lambda_{22}\alpha_{1,r} & \lambda_{12}\alpha_{1,r}-\lambda_{11}\alpha_{2,r} & 0 & 0 & \lambda_{11}\lambda_{22}-\lambda_{12}\lambda_{21} \end{bmatrix},$$

where $\omega = -(\lambda_{11}+\lambda_{22}+\phi_1\lambda_{12}+\phi_2\lambda_{21})$. Let $A_3^* = [(a_{1,3}+\rho_1 a_{4,3})', (a_{2,3}+\rho_2 a_{4,3})', (a_{3,3}+\rho_3 a_{4,3})']'$, $B_3^* = [(b_{1,3}+\rho_1 b_{4,3})', (b_{2,3}+\rho_2 b_{4,3})', (b_{3,3}+\rho_3 b_{4,3})']'$, and $C_3^* = [(c_{1,3}+\eta_1 c_{3,3})', (c_{2,3}+\eta_2 c_{3,3})']'$. Let $A_3^{**} = [(a_{1,3}+\eta_1 a_{3,3}+\eta_1\eta_2 a_{4,3})', (a_{2,3}+\eta_2 a_{3,3}+\eta_1 a_{4,3}+\eta_2^2 a_{4,3})']'$ and $B_3^{**} = [(b_{1,3}+\eta_1 b_{3,3}+$

$\eta_1\eta_2 b_{4,3})', (b_{2,3}+\eta_2 b_{3,3}+\eta_1 b_{4,3}+\eta_2^2 b_{4,3})']'$. Then,

$$D_3 = [A_3', B_3', C_3']' \quad (51)$$

$$D_3^* = [A_3^{*\prime}, B_3^{*\prime}, C_3']' \quad (52)$$

$$D_3^\dagger = [A_3', B_3', C_3^{*\prime}]' \quad (53)$$

$$D_3^\ddagger = [A_3^{*\prime}, B_3^{*\prime}, C_3^{*\prime}]' \quad (54)$$

$$D_3^\sharp = [A_3^{**\prime}, B_3^{**\prime}, C_3^{*\prime}]'. \quad (55)$$

C Lemmas

In this appendix, we provide some useful lemmas for the proofs of the asymptotic properties of the proposed estimators. To simplify notations, we drop the K subscript on Q_K and P_K. Let $\|A\| = \sqrt{\mathrm{tr}(A'A)}$ denote the Frobenius (Euclidean) norm for an $m \times n$ matrix A unless noted otherwise.

Lemma C.1 *Under Assumption 5, there exist matrices π_1 and π_2 such that, for $i = 1, 2$, $\frac{1}{n}||F_i - Q_K\pi_i||^2 \to 0$ as $n, K \to \infty$.*

Proof. See Liu and Lee (2010). ■

Lemma C.2 (i) $\mathrm{tr}(P) = K$. (ii) *Suppose that $\{A\}$ is a sequence of $n\times n$ UB matrices. For $B = PA$, $\mathrm{tr}(B) = O(K)$, $\mathrm{tr}(B^2) = O(K)$, and $\sum_i B_{ii}^2 = O(K)$, where B_{ii}'s are diagonal elements of B.*

Proof. See Liu and Lee (2010). ■

Lemma C.3 *Let $\{A\}$ and $\{B\}$ be sequences of $n \times n$ UB matrices. For $i, j = 1, 2$,* (i) $\Delta_{ij} = \frac{1}{n}\mathrm{tr}(F_i'(I-P)F_j) = o(1)$; (ii) $\frac{1}{n}F_i'PA\epsilon_j = O_p(\sqrt{K/n})$; (iii) $\frac{1}{n}\epsilon_i'B'PA\epsilon_j = O_p(K/n)$; (iv) $\frac{1}{\sqrt{n}}F_i'(I-P)A\epsilon_j = O_p(\sqrt{\Delta_{ii}})$; *and* (v) $\frac{1}{\sqrt{n}}[\epsilon_i'PA\epsilon_j - \mathrm{E}(\epsilon_i'PA\epsilon_j)] = O_p(\sqrt{K/n})$.

Proof. For (i), $\frac{1}{n}\mathrm{tr}(F_i'(I-P)F_j) = \frac{1}{n}\mathrm{tr}((F_i-Q\pi_i)'(I-P)(F_j-Q\pi_j)) \le [\frac{1}{n}\mathrm{tr}((F_i-Q\pi_i)'(F_i-Q\pi_i))]^{1/2}[\frac{1}{n}\mathrm{tr}((F_j-Q\pi_j)'(I-P)(F_j-Q\pi_j))]^{1/2} \le [\frac{1}{n}\mathrm{tr}((F_i-Q\pi_i)'(F_i-Q\pi_i))]^{1/2}[\frac{1}{n}\mathrm{tr}((F_j-Q\pi_j)'(F_j-Q\pi_j))]^{1/2} = [\frac{1}{n}||F_i-Q\pi_i||^2]^{1/2}[\frac{1}{n}||F_j-Q\pi_j||^2]^{1/2} \to 0$, by Assumption 5 and Lemma C.1. For (ii) and (iii), see Liu and Lee (2010). For (iv), as $\mathrm{Var}[\frac{1}{\sqrt{n}}F_i'(I-P)A\epsilon_j] = \frac{\sigma_j^2}{n}F_i'(I-$

$P)AA(I-P)F_i \leq \frac{\sigma_j^2}{n}F_i'(I-P)F_i||AA||_\infty = O(\Delta_{ii})$, we have $\frac{1}{\sqrt{n}}F_i'(I-P)A\epsilon_j = O_p(\sqrt{\Delta_{ii}})$ by Markov's inequality. For (v), let $M = PA$. $\mathrm{E}(\epsilon_j'M'\epsilon_i\epsilon_i'M\epsilon_j) - \mathrm{E}(\epsilon_j'M'\epsilon_i)\mathrm{E}(\epsilon_i'M\epsilon_j) = c_1\sum_s M_{ss}^2 + c_2\mathrm{tr}(M^2) + c_3\mathrm{tr}(M'M)$, where c_1, c_2, c_3 are functions of moments of ϵ_i, ϵ_j and they are bounded by finite constants by Assumption 1. As $\sum_s M_{ss}^2, \mathrm{tr}(M^2), \mathrm{tr}(M'M)$ are $O(K)$ by Lemma C.2, we have $\frac{1}{\sqrt{n}}[\epsilon_i'PA\epsilon_j - \mathrm{E}(\epsilon_i'PA\epsilon_j)] = O_p(\sqrt{K/n})$ by Markov's inequality. ■

D Proofs

Proof of Proposition 1. $\mathrm{E}(JZ_1) = J[\mathrm{E}(GY_1), X, GX]$ has full column rank if, for some r,

$$J_r[\mathrm{E}(G_rY_{1,r})d_1 + X_rd_2 + G_rX_rd_3] = 0 \tag{56}$$

implies $d_1 = d_2 = d_3 = 0$. As $J_r = I_{n_r} - \frac{1}{n_r}l_{n_r}l_{n_r}'$, (56) can be rewritten as

$$\mathrm{E}(G_rY_{1,r})d_1 + X_rd_2 + G_rX_rd_3 + l_{n_r}\mu = 0, \tag{57}$$

where $\mu = -\frac{1}{n_r}l_{n_r}'[\mathrm{E}(G_rY_{1,r})d_1 + X_rd_2 + G_rX_rd_3]$. Premultiply (57) by $(I_{n_r} - \lambda_{11}G_r)$. As

$$(I_{n_r} - \lambda_{11}G_r)\mathrm{E}(G_rY_{1,r}) = G_rX_r\beta_1 + G_r^2X_r\gamma_1 + G_rl_{n_r}\alpha_{1,r}$$

from the reduced-form equation, we have

$$X_rd_2 + G_rX_r(\beta_1d_1 - \lambda_{11}d_2 + d_3) + G_r^2X_r(\gamma_1d_1 - \lambda_{11}d_3) + l_{n_r}\mu + G_rl_{n_r}(\alpha_{1,r}d_1 - \lambda_{11}\mu) = 0.$$

Suppose G_r has non-constant row sums. We consider 2 cases. (i) I_{n_r}, G_r, G_r^2 are linearly independent. In this case, $d_2 = \beta_1d_1 - \lambda_{11}d_2 + d_3 = \gamma_1d_1 - \lambda_{11}d_3 = \mu = \alpha_{1,r}d_1 - \lambda_{11}\mu = 0$, which implies $d_1 = d_2 = d_3 = \mu = 0$ if $\alpha_{1,r} \neq 0$ or $\lambda_{11}\beta_1 + \gamma_1 \neq 0$. (ii) $G_r^2 = \rho_1I_{n_r} + \rho_2G_r$. In this case, $d_2 + \rho_1(\gamma_1d_1 - \lambda_{11}d_3) = \beta_1d_1 - \lambda_{11}d_2 + d_3 + \rho_2(\gamma_1d_1 - \lambda_{11}d_3) = \mu = \alpha_{1,r}d_1 - \lambda_{11}\mu = 0$, which implies $d_1 = d_2 = d_3 = \mu = 0$ if $\alpha_{1,r} \neq 0$ and $1 - \rho_2\lambda_{11} - \rho_1\lambda_{11}^2 \neq 0$. ■

Proof of Corollary 1. For a symmetric adjacency matrix G, I, G, G^2 are linearly independent if G has non-constant row sums. This can be shown by contradiction. As elements of G are either one or zero, the ith diagonal element of G^2 equals $\sum_j g_{ij}g_{ji} = \sum_j g_{ij}^2 = \sum_j g_{ij}$. Therefore, if I, G, G^2 are linearly dependent such that $G^2 = \rho_1I + \rho_2G$, then all the diagonal elements of G^2 equal ρ_1, i.e.,

$\sum_j g_{ij} = \rho_1$ for all i. This is a contradiction as G has non-constant row sums. The desired result follows from Proposition 1(i). ■

Proof of Proposition 2. $\mathrm{E}(JZ_1) = J[\mathrm{E}(Y_2), \mathrm{E}(GY_1), X, GX]$ has full column rank if, for some r,

$$J_r[\mathrm{E}(Y_{2,r})d_1 + \mathrm{E}(G_rY_{1,r})d_2 + X_rd_3 + G_rX_rd_4] = 0 \tag{58}$$

implies $d_1 = d_2 = d_3 = d_4 = 0$. As $J_r = I_{n_r} - \frac{1}{n_r}l_{n_r}l'_{n_r}$, (58) can be rewritten as

$$\mathrm{E}(Y_{2,r})d_1 + \mathrm{E}(G_rY_{1,r})d_2 + X_rd_3 + G_rX_rd_4 + l_{n_r}\mu = 0, \tag{59}$$

where $\mu = -\frac{1}{n_r}l'_{n_r}[\mathrm{E}(Y_{2,r})d_1 + \mathrm{E}(G_rY_{1,r})d_2 + X_rd_3 + G_rX_rd_4]$. Under the exclusion restrictions $\lambda_{21} = \lambda_{12} = 0$, the reduced-form equations (16) and (17) become

$$\begin{aligned} S_{\lambda,r}\mathrm{E}(Y_{1,r}) &= X_r(\phi_1\beta_2 + \beta_1) + G_rX_r(\gamma_1 + \phi_1\gamma_2 - \lambda_{22}\beta_1) - G_r^2X_r\lambda_{22}\gamma_1 \\ &\quad + l_{n_r}(\phi_1\alpha_{2,r} + \alpha_{1,r}) - G_rl_{n_r}\lambda_{22}\alpha_{1,r} \end{aligned} \tag{60}$$

$$\begin{aligned} S_{\lambda,r}\mathrm{E}(Y_{2,r}) &= X_r(\phi_2\beta_1 + \beta_2) + G_rX_r(\gamma_2 + \phi_2\gamma_1 - \lambda_{11}\beta_2) - G_r^2X_r\lambda_{11}\gamma_2 \\ &\quad + l_{n_r}(\phi_2\alpha_{1,r} + \alpha_{2,r}) - G_rl_{n_r}\lambda_{11}\alpha_{2,r} \end{aligned} \tag{61}$$

where $S_{\lambda,r} = (1 - \phi_1\phi_2)I_{n_r} - (\lambda_{11} + \lambda_{22})G_r + \lambda_{11}\lambda_{22}G_r^2$. Premultiply (59) by $S_{\lambda,r}$. As $G_rS_{\lambda,r} = S_{\lambda,r}G_r$, it follows from (60) and (61) that

$$X_ra_1 + G_rX_ra_2 + G_r^2X_ra_3 + G_r^3X_ra_4 + l_{n_r}c_1 + G_rl_{n_r}c_2 + G_r^2l_{n_r}c_3 = 0,$$

where

$$\begin{aligned} a_1 &= (\phi_2\beta_1 + \beta_2)d_1 + (1 - \phi_1\phi_2)d_3 \\ a_2 &= (\phi_2\gamma_1 + \gamma_2 - \lambda_{11}\beta_2)d_1 + (\beta_1 + \phi_1\beta_2)d_2 - (\lambda_{11} + \lambda_{22})d_3 + (1 - \phi_1\phi_2)d_4 \\ a_3 &= -\lambda_{11}\gamma_2d_1 + (\gamma_1 + \phi_1\gamma_2 - \lambda_{22}\beta_1)d_2 + \lambda_{11}\lambda_{22}d_3 - (\lambda_{11} + \lambda_{22})d_4 \\ a_4 &= -\lambda_{22}\gamma_1d_2 + \lambda_{11}\lambda_{22}d_4 \end{aligned}$$

and

$$\begin{aligned} c_1 &= (\phi_2\alpha_{1,r} + \alpha_{2,r})d_1 + (1-\phi_1\phi_2)\mu \\ c_2 &= -\lambda_{11}\alpha_{2,r}d_1 + (\alpha_{1,r} + \phi_1\alpha_{2,r})d_2 - (\lambda_{11}+\lambda_{22})\mu \\ c_3 &= -\lambda_{22}\alpha_{1,r}d_2 + \lambda_{11}\lambda_{22}\mu. \end{aligned}$$

Suppose G_r has non-constant row sums and I_{n_r}, G_r, G_r^2 are linearly independent. First, we consider the case that $[l_{n_r}, G_r l_{n_r}, G_r^2 l_{n_r}]$ has full column rank. In this case, if $I_{n_r}, G_r, G_r^2, G_r^3$ are linearly independent, then $a_1 = a_2 = a_3 = a_4 = c_1 = c_2 = c_3 = 0$, which implies $d_1 = d_2 = d_3 = d_4 = \mu = 0$ if D_1 given by (43) has full rank. If $G_r^3 = \rho_1 I_{n_r} + \rho_2 G_r + \rho_3 G_r^2$, then $a_1 + \rho_1 a_4 = a_2 + \rho_2 a_4 = a_3 + \rho_3 a_4 = c_1 = c_2 = c_3 = 0$, which implies $d_1 = d_2 = d_3 = d_4 = \mu = 0$ if D_1^* given by (44) has full rank.

Next, we consider the case that $G_r^2 l_{n_r} = \eta_1 l_{n_r} + \eta_2 G_r l_{n_r}$. In this case, if $I_{n_r}, G_r, G_r^2, G_r^3$ are linearly independent, then $a_1 = a_2 = a_3 = a_4 = c_1 + \eta_1 c_3 = c_2 + \eta_2 c_3 = 0$, which implies $d_1 = d_2 = d_3 = d_4 = \mu = 0$ if $D_1^\dagger$ given by (45) has full rank. If $G_r^3 = \rho_1 I_{n_r} + \rho_2 G_r + \rho_3 G_r^2$, then $a_1 + \rho_1 a_4 = a_2 + \rho_2 a_4 = a_3 + \rho_3 a_4 = c_1 + \eta_1 c_3 = c_2 + \eta_2 c_3 = 0$, which implies $d_1 = d_2 = d_3 = d_4 = \mu = 0$ if $D_1^\ddagger$ given by (46) has full rank. ■

Proof of Proposition 3. $\mathrm{E}(JZ_1) = J[\mathrm{E}(GY_1), \mathrm{E}(GY_2), X, GX]$ has full column rank if, for some r,

$$J_r[\mathrm{E}(G_rY_{1,r})d_1 + \mathrm{E}(G_rY_{2,r})d_2 + X_rd_3 + G_rX_rd_4] = 0 \tag{62}$$

implies $d_1 = d_2 = d_3 = d_4 = 0$. As $J_r = I_{n_r} - \frac{1}{n_r} l_{n_r} l_{n_r}'$, (62) can be rewritten as

$$\mathrm{E}(G_rY_{1,r})d_1 + \mathrm{E}(G_rY_{2,r})d_2 + X_rd_3 + G_rX_rd_4 + l_{n_r}\mu = 0, \tag{63}$$

where $\mu = -\frac{1}{n_r} l_{n_r}'[\mathrm{E}(G_rY_{1,r})d_1 + \mathrm{E}(G_rY_{2,r})d_2 + X_rd_3 + G_rX_rd_4]$. Under the exclusion restrictions $\phi_1 = \phi_2 = 0$, the reduced-form equations (16) and (17) become

$$\begin{aligned} S_{\phi,r}\mathrm{E}(Y_{1,r}) &= X_r\beta_1 + G_rX_r(\lambda_{21}\beta_2 - \lambda_{22}\beta_1 + \gamma_1) + G_r^2X_r(\lambda_{21}\gamma_2 - \lambda_{22}\gamma_1) \\ &\quad + l_{n_r}\alpha_{1,r} + G_r l_{n_r}(\lambda_{21}\alpha_{2,r} - \lambda_{22}\alpha_{1,r}) \end{aligned} \tag{64}$$

$$\begin{aligned} S_{\phi,r}\mathrm{E}(Y_{2,r}) &= X_r\beta_2 + G_rX_r(\lambda_{12}\beta_1 - \lambda_{11}\beta_2 + \gamma_2) + G_r^2X_r(\lambda_{12}\gamma_1 - \lambda_{11}\gamma_2) \\ &\quad + l_{n_r}\alpha_{2,r} + G_rl_{n_r}(\lambda_{12}\alpha_{1,r} - \lambda_{11}\alpha_{2,r}) \end{aligned} \tag{65}$$

where $S_{\phi,r} = I_{n_r} - (\lambda_{11} + \lambda_{22})G_r + (\lambda_{11}\lambda_{22} - \lambda_{12}\lambda_{21})G_r^2$. Premultiply (63) by $S_{\phi,r}$. As $G_rS_{\phi,r} = S_{\phi,r}G_r$, it follows from (64) and (65) that

$$X_ra_1 + G_rX_ra_2 + G_r^2X_ra_3 + G_r^3X_ra_4 + l_{n_r}c_1 + G_rl_{n_r}c_2 + G_r^2l_{n_r}c_3 = 0,$$

where

$$\begin{aligned} a_1 &= d_3 \\ a_2 &= \beta_1d_1 + \beta_2d_2 - (\lambda_{11} + \lambda_{22})d_3 + d_4 \\ a_3 &= (\lambda_{21}\beta_2 - \lambda_{22}\beta_1 + \gamma_1)d_1 + (\lambda_{12}\beta_1 - \lambda_{11}\beta_2 + \gamma_2)d_2 + (\lambda_{11}\lambda_{22} - \lambda_{12}\lambda_{21})d_3 - (\lambda_{11} + \lambda_{22})d_4 \\ a_4 &= (\lambda_{21}\gamma_2 - \lambda_{22}\gamma_1)d_1 + (\lambda_{12}\gamma_1 - \lambda_{11}\gamma_2)d_2 + (\lambda_{11}\lambda_{22} - \lambda_{12}\lambda_{21})d_4 \end{aligned}$$

and

$$\begin{aligned} c_1 &= \mu \\ c_2 &= \alpha_{1,r}d_1 + \alpha_{2,r}d_2 - (\lambda_{11} + \lambda_{22})\mu \\ c_3 &= (\lambda_{21}\alpha_{2,r} - \lambda_{22}\alpha_{1,r})d_1 + (\lambda_{12}\alpha_{1,r} - \lambda_{11}\alpha_{2,r})d_2 + (\lambda_{11}\lambda_{22} - \lambda_{12}\lambda_{21})\mu. \end{aligned}$$

Suppose G_r has non-constant row sums and I_{n_r}, G_r, G_r^2 are linearly independent. First, we consider the case that $[l_{n_r}, G_rl_{n_r}, G_r^2l_{n_r}]$ has full column rank. In this case, if $I_{n_r}, G_r, G_r^2, G_r^3$ are linearly independent, then $a_1 = a_2 = a_3 = a_4 = c_1 = c_2 = c_3 = 0$, which implies $d_1 = d_2 = d_3 = d_4 = \mu = 0$ if D_2 given by (47) has full rank. If $G_r^3 = \rho_1I_{n_r} + \rho_2G_r + \rho_3G_r^2$, then $a_1 + \rho_1a_4 = a_2 + \rho_2a_4 = a_3 + \rho_3a_4 = c_1 = c_2 = c_3 = 0$, which implies $d_1 = d_2 = d_3 = d_4 = \mu = 0$ if D_2^* given by (48) has full rank.

Next, we consider the case that $G_r^2l_{n_r} = \eta_1l_{n_r} + \eta_2G_rl_{n_r}$. In this case, if $I_{n_r}, G_r, G_r^2, G_r^3$ are linearly independent, then $a_1 = a_2 = a_3 = a_4 = c_1 + \eta_1c_3 = c_2 + \eta_2c_3 = 0$, which implies $d_1 = d_2 = d_3 = d_4 = \mu = 0$ if $D_2^\dagger$ given by (49) has full rank. If $G_r^3 = \rho_1I_{n_r} + \rho_2G_r + \rho_3G_r^2$, then $a_1 + \rho_1a_4 = a_2 + \rho_2a_4 = a_3 + \rho_3a_4 = c_1 + \eta_1c_3 = c_2 + \eta_2c_3 = 0$, which implies $d_1 = d_2 = d_3 = d_4 =$

$\mu = 0$ if $D_2^{\ddagger}$ given by (50) has full rank. ■

Proof of Proposition 4. $\mathrm{E}(\bar{Z}_1) = [\mathrm{E}(Y_2), \mathrm{E}(GY_1), \mathrm{E}(GY_2), X, GX, L]$ has full column rank if and only if

$$\mathrm{E}(Y_2)d_1 + \mathrm{E}(GY_1)d_2 + \mathrm{E}(GY_2)d_3 + Xd_4 + GXd_5 + Ld_6 = 0 \tag{66}$$

implies $d_1 = d_2 = d_3 = d_4 = d_5 = d_6 = 0$. Premultiply (66) by S. As $GS = SG$, it follows from (16) and (17) that

$$Xa_1 + GXa_2 + G^2Xa_3 + G^3Xa_4 + Lc_1 + GLc_2 + G^2Lc_3 = 0,$$

where

$$\begin{aligned}
a_1 &= (\phi_2\beta_1 + \beta_2)d_1 + (1 - \phi_1\phi_2)d_4 \\
a_2 &= (\lambda_{12}\beta_1 - \lambda_{11}\beta_2 + \phi_2\gamma_1 + \gamma_2)d_1 + (\phi_1\beta_2 + \beta_1)d_2 + (\phi_2\beta_1 + \beta_2)d_3 \\
&\quad -(\lambda_{11} + \lambda_{22} + \phi_1\lambda_{12} + \phi_2\lambda_{21})d_4 + (1 - \phi_1\phi_2)d_5 \\
a_3 &= (\lambda_{12}\gamma_1 - \lambda_{11}\gamma_2)d_1 + (\lambda_{21}\beta_2 - \lambda_{22}\beta_1 + \phi_1\gamma_2 + \gamma_1)d_2 + (\lambda_{12}\beta_1 - \lambda_{11}\beta_2 + \phi_2\gamma_1 + \gamma_2)d_3 \\
&\quad +(\lambda_{11}\lambda_{22} - \lambda_{12}\lambda_{21})d_4 - (\lambda_{11} + \lambda_{22} + \phi_1\lambda_{12} + \phi_2\lambda_{21})d_5 \\
a_4 &= (\lambda_{21}\gamma_2 - \lambda_{22}\gamma_1)d_2 + (\lambda_{12}\gamma_1 - \lambda_{11}\gamma_2)d_3 + (\lambda_{11}\lambda_{22} - \lambda_{12}\lambda_{21})d_5
\end{aligned}$$

and

$$\begin{aligned}
c_1 &= (\phi_2\alpha_1 + \alpha_2)d_1 + (1 - \phi_1\phi_2)d_6 \\
c_2 &= (\lambda_{12}\alpha_1 - \lambda_{11}\alpha_2)d_1 + (\phi_1\alpha_2 + \alpha_1)d_2 + (\phi_2\alpha_1 + \alpha_2)d_3 - (\lambda_{11} + \lambda_{22} + \phi_1\lambda_{12} + \phi_2\lambda_{21})d_6 \\
c_3 &= (\lambda_{21}\alpha_2 - \lambda_{22}\alpha_1)d_2 + (\lambda_{12}\alpha_1 - \lambda_{11}\alpha_2)d_3 + (\lambda_{11}\lambda_{22} - \lambda_{12}\lambda_{21})d_6.
\end{aligned}$$

Then, if $[X, GX, G^2X, G^3X, L, GL, G^2L]$ has full column rank, we have $a_1 = a_2 = a_3 = a_4 = c_1 = c_2 = c_3 = 0$, which implies $d_2 = (\lambda_{12} + \phi_2\lambda_{11})d_1/(\phi_1\phi_2 - 1)$, $d_3 = (\lambda_{22} + \phi_2\lambda_{21})d_1/(\phi_1\phi_2 - 1)$, $d_4 = (\beta_2 + \phi_2\beta_1)d_1/(\phi_1\phi_2 - 1)$, $d_5 = (\gamma_2 + \phi_2\gamma_1)d_1/(\phi_1\phi_2 - 1)$ and $d_6 = (\alpha_2 + \phi_2\alpha_1)d_1/(\phi_1\phi_2 - 1)$. Therefore, $\mathrm{E}(\bar{Z}_1)$ does not have full column rank. We can show that $\mathrm{E}(\bar{Z}_2)$ does not have full column rank by the same token. ■

Proof of Proposition 5. $\mathrm{E}(JZ_1) = J[\mathrm{E}(Y_2), \mathrm{E}(GY_1), \mathrm{E}(GY_2), X_1, GX_1]$ has full column rank if, for some r,

$$J_r[\mathrm{E}(Y_{2,r})d_1 + \mathrm{E}(G_rY_{1,r})d_2 + \mathrm{E}(G_rY_{2,r})d_3 + X_{1,r}d_4 + G_rX_{1,r}d_5] = 0 \tag{67}$$

implies $d_1 = d_2 = d_3 = d_4 = d_5 = 0$. As $J_r = I_{n_r} - \frac{1}{n_r}l_{n_r}l'_{n_r}$, (67) can be rewritten as

$$\mathrm{E}(Y_{2,r})d_1 + \mathrm{E}(G_rY_{1,r})d_2 + \mathrm{E}(G_rY_{2,r})d_3 + X_{1,r}d_4 + G_rX_{1,r}d_5 + l_{n_r}\mu = 0, \tag{68}$$

where $\mu = -\frac{1}{n_r}l'_{n_r}[\mathrm{E}(Y_{2,r})d_1 + \mathrm{E}(G_rY_{1,r})d_2 + \mathrm{E}(G_rY_{2,r})d_3 + X_{1,r}d_4 + G_rX_{1,r}d_5]$. Premultiply (68) by S_r. As $G_rS_r = S_rG_r$, it follows from (31) and (32) that

$$\begin{aligned} 0 &= X_{1,r}a_1 + G_rX_{1,r}a_2 + G_r^2X_{1,r}a_3 + G_r^3X_{1,r}a_4 + X_{2,r}b_1 + G_rX_{2,r}b_2 + G_r^2X_{2,r}b_3 + G_r^3X_{2,r}b_4 \\ &\quad + l_{n_r}c_1 + G_rl_{n_r}c_2 + G_r^2l_{n_r}c_3, \end{aligned}$$

where

$$\begin{aligned} a_1 &= \phi_2\beta_1d_1 + (1-\phi_1\phi_2)d_4 \\ a_2 &= (\lambda_{12}\beta_1 + \phi_2\gamma_1)d_1 + \beta_1d_2 + \phi_2\beta_1d_3 - (\lambda_{11} + \lambda_{22} + \phi_1\lambda_{12} + \phi_2\lambda_{21})d_4 + (1-\phi_1\phi_2)d_5 \\ a_3 &= \lambda_{12}\gamma_1d_1 + (\gamma_1 - \lambda_{22}\beta_1)d_2 + (\lambda_{12}\beta_1 + \phi_2\gamma_1)d_3 + (\lambda_{11}\lambda_{22} - \lambda_{12}\lambda_{21})d_4 - (\lambda_{11} + \lambda_{22} + \phi_1\lambda_{12} + \phi_2\lambda_{21})d_5 \\ a_4 &= -\lambda_{22}\gamma_1d_2 + \lambda_{12}\gamma_1d_3 + (\lambda_{11}\lambda_{22} - \lambda_{12}\lambda_{21})d_5 \end{aligned}$$

$$\begin{aligned} b_1 &= \beta_2d_1 \\ b_2 &= (\gamma_2 - \lambda_{11}\beta_2)d_1 + \phi_1\beta_2d_2 + \beta_2d_3 \\ b_3 &= -\lambda_{11}\gamma_2d_1 + (\lambda_{21}\beta_2 + \phi_1\gamma_2)d_2 + (\gamma_2 - \lambda_{11}\beta_2)d_3 \\ b_4 &= \lambda_{21}\gamma_2d_2 - \lambda_{11}\gamma_2d_3 \end{aligned}$$

and

$$c_1 = (\alpha_{2,r} + \phi_2\alpha_{1,r})d_1 + (1-\phi_1\phi_2)\mu$$

$$c_2 = (\lambda_{12}\alpha_{1,r} - \lambda_{11}\alpha_{2,r})d_1 + (\alpha_{1,r} + \phi_1\alpha_{2,r})d_2 + (\alpha_{2,r} + \phi_2\alpha_{1,r})d_3 - (\lambda_{11} + \lambda_{22} + \phi_1\lambda_{12} + \phi_2\lambda_{21})\mu$$

$$c_3 = (\lambda_{21}\alpha_{2,r} - \lambda_{22}\alpha_{1,r})d_2 + (\lambda_{12}\alpha_{1,r} - \lambda_{11}\alpha_{2,r})d_3 + (\lambda_{11}\lambda_{22} - \lambda_{12}\lambda_{21})\mu.$$

Suppose G_r has non-constant row sums, $X_{1,r}, X_{2,r}$ are vectors and $[X_{1,r}, X_{2,r}]$ has full column rank. First, we consider the case that $[l_{n_r}, G_r l_{n_r}, G_r^2 l_{n_r}]$ has full column rank. In this case, if $I_{n_r}, G_r, G_r^2, G_r^3$ are linearly independent, then $a_1 = a_2 = a_3 = a_4 = b_1 = b_2 = b_3 = b_4 = c_1 = c_2 = c_3 = 0$, which implies $d_1 = d_2 = d_3 = d_4 = d_5 = \mu = 0$ if D_3 given by (51) has full rank. If I_{n_r}, G_r, G_r^2 are linearly independent and $G_r^3 = \rho_1 I_{n_r} + \rho_2 G_r + \rho_3 G_r^2$, then $a_1 + \rho_1 a_4 = a_2 + \rho_2 a_4 = a_3 + \rho_3 a_4 = b_1 + \rho_1 b_4 = b_2 + \rho_2 b_4 = b_3 + \rho_3 b_4 = c_1 = c_2 = c_3 = 0$, which implies $d_1 = d_2 = d_3 = d_4 = d_5 = \mu = 0$ if D_3^* given by (52) has full rank.

Next, we consider the case that $G_r^2 l_{n_r} = \eta_1 l_{n_r} + \eta_2 G_r l_{n_r}$. In this case, if $I_{n_r}, G_r, G_r^2, G_r^3$ are linearly independent, then $a_1 = a_2 = a_3 = a_4 = b_1 = b_2 = b_3 = b_4 = c_1 + \eta_1 c_3 = c_2 + \eta_2 c_3 = 0$, which implies $d_1 = d_2 = d_3 = d_4 = d_5 = \mu = 0$ if $D_3^\dagger$ given by (53) has full rank. If I_{n_r}, G_r, G_r^2 are linearly independent and $G_r^3 = \rho_1 I_{n_r} + \rho_2 G_r + \rho_3 G_r^2$, then $a_1 + \rho_1 a_4 = a_2 + \rho_2 a_4 = a_3 + \rho_3 a_4 = b_1 + \rho_1 b_4 = b_2 + \rho_2 b_4 = b_3 + \rho_3 b_4 = c_1 + \eta_1 c_3 = c_2 + \eta_2 c_3 = 0$, which implies $d_1 = d_2 = d_3 = d_4 = d_5 = \mu = 0$ if $D_3^\ddagger$ given by (54) has full rank. Finally, if $G_r^2 = \eta_1 I_{n_r} + \eta_2 G_r$, then $G_r^3 = \eta_1\eta_2 I_{n_r} + (\eta_1 + \eta_2^2)G_r$, then $a_1 + \eta_1 a_3 + \eta_1\eta_2 a_4 = a_2 + \eta_2 a_3 + (\eta_1 + \eta_2^2)a_4 = b_1 + \eta_1 b_3 + \eta_1\eta_2 b_4 = b_2 + \eta_2 b_3 + (\eta_1 + \eta_2^2)b_4 = c_1 + \eta_1 c_3 = c_2 + \eta_2 c_3 = 0$, which implies $d_1 = d_2 = d_3 = d_4 = d_5 = \mu = 0$ if $D_3^\sharp$ given by (55) has full rank. ■

Proof of Proposition 6. In this proof, we focus on $\hat{\delta}_{1,2sls}$. The results for $\hat{\delta}_{2,2sls}$ can be derived by the same argument. $\sqrt{n}(\hat{\delta}_{1,2sls} - \delta_1 - b_{1,2sls}) = (\frac{1}{n}Z_1'PZ_1)^{-1}\frac{1}{\sqrt{n}}[Z_1'P\epsilon_1 - \mathrm{E}(U_1'P\epsilon_1)]$. As $JZ_1 = F_1 + U_1$, we have $\frac{1}{n}Z_1'PZ_1 = \frac{1}{n}F_1'F_1 - \frac{1}{n}F_1'(I-P)F_1 + \frac{1}{n}F_1'PU_1 + \frac{1}{n}U_1'PF_1 + \frac{1}{n}U_1'PU_1$ and $\frac{1}{\sqrt{n}}[Z_1'P\epsilon_1 - \mathrm{E}(U_1'P\epsilon_1)] = \frac{1}{\sqrt{n}}F_1'\epsilon_1 - \frac{1}{\sqrt{n}}F_1'(I-P)\epsilon_1 + \frac{1}{\sqrt{n}}[U_1'P\epsilon_1 - \mathrm{E}(U_1'P\epsilon_1)]$. As $K/n \to 0$ and $U_1 = J[S^{-1}u_2, GS^{-1}u_1, GS^{-1}u_2, 0]$, where $u_1 = (I - \lambda_{22}G)\epsilon_1 + (\phi_1 I + \lambda_{21}G)\epsilon_2$ and $u_2 = (I - \lambda_{11}G)\epsilon_2 + (\phi_2 I + \lambda_{12}G)\epsilon_1$, it follows by Lemma C.3 that $\frac{1}{n}Z_1'PZ_1 = H_{11} + O(\Delta_{11}) + O_p(\sqrt{K/n}) = H_{11} + o_p(1)$ and $\frac{1}{\sqrt{n}}[Z_1'P\epsilon_1 - \mathrm{E}(U_1'P\epsilon_1)] = \frac{1}{\sqrt{n}}F_1'\epsilon_1 + O_p(\sqrt{\Delta_{11}}) + O_p(\sqrt{K/n}) = \frac{1}{\sqrt{n}}F_1'\epsilon_1 + o_p(1)$. As $\frac{1}{\sqrt{n}}F_1'\epsilon_1 \xrightarrow{d} N(0, \sigma_1^2 H_{11})$ by Theorem A in Kelejian and Prucha (1999), the asymptotic distribution of $\hat{\delta}_{1,2sls}$ follows by Slutsky's theorem. Furthermore, as $\frac{1}{n}\mathrm{E}(U_1'P\epsilon_1)$ given by (36) is $O(K/n)$ by Lemma C.2, $b_{1,2sls} = O_p(K/n)$. ■

Proof of Proposition 7. From the proof of Proposition 6, it is sufficient to show that $\frac{1}{\sqrt{n}}[\hat{\mathrm{E}}(U_1'P\epsilon_1) - \mathrm{E}(U_1'P\epsilon_1)] = o_p(1)$ and $\frac{1}{\sqrt{n}}[\hat{\mathrm{E}}(U_2'P\epsilon_2) - \mathrm{E}(U_2'P\epsilon_2)] = o_p(1)$. Here, we show the first element of $\frac{1}{\sqrt{n}}[\hat{\mathrm{E}}(U_1'P\epsilon_1) - \mathrm{E}(U_1'P\epsilon_1)]$, i.e. $\frac{1}{\sqrt{n}}[(\tilde{\sigma}_{12} + \tilde{\phi}_2\tilde{\sigma}_1^2)\mathrm{tr}(P\tilde{S}^{-1}) - (\sigma_{12} + \phi_2\sigma_1^2)\mathrm{tr}(PS^{-1})] + \frac{1}{\sqrt{n}}[(\tilde{\lambda}_{12}\tilde{\sigma}_1^2 - \tilde{\lambda}_{11}\tilde{\sigma}_{12})\mathrm{tr}(P\tilde{S}^{-1}G) - (\lambda_{12}\sigma_1^2 - \lambda_{11}\sigma_{12})\mathrm{tr}(PS^{-1}G)]$, is $o_p(1)$, where $\tilde{S} = (1 - \tilde{\phi}_1\tilde{\phi}_2)I - (\tilde{\lambda}_{11} + \tilde{\lambda}_{22} + \tilde{\phi}_1\tilde{\lambda}_{12} + \tilde{\phi}_2\tilde{\lambda}_{21})G + (\tilde{\lambda}_{11}\tilde{\lambda}_{22} - \tilde{\lambda}_{12}\tilde{\lambda}_{21})G^2$. Convergence of other terms in $\frac{1}{\sqrt{n}}[\hat{\mathrm{E}}(U_1'P\epsilon_1) - \mathrm{E}(U_1'P\epsilon_1)]$ and $\frac{1}{\sqrt{n}}[\hat{\mathrm{E}}(U_2'P\epsilon_2) - \mathrm{E}(U_2'P\epsilon_2)]$ follows a similar argument. As $\tilde{\delta}_1, \tilde{\delta}_2, \tilde{\sigma}_1^2, \tilde{\sigma}_2^2, \tilde{\sigma}_{12}$ are $\sqrt{n}$-consistent estimators and $\sqrt{n}(\tilde{S}^{-1} - S^{-1}) = \sqrt{n}S^{-1}(S - \tilde{S})\tilde{S}^{-1} = \sqrt{n}(\tilde{\phi}_1\tilde{\phi}_2 - \phi_1\phi_2)S^{-1}\tilde{S}^{-1} + [\sqrt{n}(\tilde{\lambda}_{11} - \lambda_{11}) + \sqrt{n}(\tilde{\lambda}_{22} - \lambda_{22}) + \sqrt{n}(\tilde{\phi}_1\tilde{\lambda}_{12} - \phi_1\lambda_{12}) + \sqrt{n}(\tilde{\phi}_2\tilde{\lambda}_{21} - \phi_2\lambda_{21})]S^{-1}G\tilde{S}^{-1} + [\sqrt{n}(\tilde{\lambda}_{12}\tilde{\lambda}_{21} - \lambda_{12}\lambda_{21}) - \sqrt{n}(\tilde{\lambda}_{11}\tilde{\lambda}_{22} - \lambda_{11}\lambda_{22})]S^{-1}G^2\tilde{S}^{-1}$, it follows that $\frac{1}{\sqrt{n}}[(\tilde{\sigma}_{12} + \tilde{\phi}_2\tilde{\sigma}_1^2)\mathrm{tr}(P\tilde{S}^{-1}) - (\sigma_{12} + \phi_2\sigma_1^2)\mathrm{tr}(PS^{-1})] = [\sqrt{n}(\tilde{\sigma}_{12} - \sigma_{12}) + \sqrt{n}(\tilde{\phi}_2\tilde{\sigma}_1^2 - \phi_2\sigma_1^2)]\frac{1}{n}\mathrm{tr}(P\tilde{S}^{-1}) + (\sigma_{12} + \phi_2\sigma_1^2)\frac{1}{n}\mathrm{tr}(P\sqrt{n}(\tilde{S}^{-1} - S^{-1})) = O_p(K/n)$ and $\frac{1}{\sqrt{n}}[(\tilde{\lambda}_{12}\tilde{\sigma}_1^2 - \tilde{\lambda}_{11}\tilde{\sigma}_{12})\mathrm{tr}(P\tilde{S}^{-1}G) - (\lambda_{12}\sigma_1^2 - \lambda_{11}\sigma_{12})\mathrm{tr}(PS^{-1}G)] = [\sqrt{n}(\tilde{\lambda}_{12}\tilde{\sigma}_1^2 - \lambda_{12}\sigma_1^2) - \sqrt{n}(\tilde{\lambda}_{11}\tilde{\sigma}_{12} - \lambda_{11}\sigma_{12})]\frac{1}{n}\mathrm{tr}(P\tilde{S}^{-1}G) + (\lambda_{12}\sigma_1^2 - \lambda_{11}\sigma_{12})\frac{1}{n}\mathrm{tr}(GP\sqrt{n}(\tilde{S}^{-1} - S^{-1})) = O_p(K/n)$. The desired result follows as $K/n \to 0$. ■

Proof of Proposition 8. First, we consider the infeasible 3SLS estimator $\tilde{\delta}_{3sls} = [Z'(\Sigma^{-1} \otimes P)Z]^{-1}Z'(\Sigma^{-1} \otimes P)Y$ such that $\sqrt{n}(\tilde{\delta}_{3sls} - \delta - b_{3sls}) = [\frac{1}{n}Z'(\Sigma^{-1} \otimes P)Z]^{-1}\frac{1}{\sqrt{n}}\{Z'(\Sigma^{-1} \otimes P)\epsilon - \mathrm{E}[U'(\Sigma^{-1} \otimes P)\epsilon]\}$. As $(I_2 \otimes J)Z = F + U$, we have $\frac{1}{n}Z'(\Sigma^{-1} \otimes P)Z = \frac{1}{n}F'(\Sigma^{-1} \otimes I)F - \frac{1}{n}F'[\Sigma^{-1} \otimes (I - P)]F + \frac{1}{n}F'(\Sigma^{-1} \otimes P)U + \frac{1}{n}U'(\Sigma^{-1} \otimes P)F + \frac{1}{n}U'(\Sigma^{-1} \otimes P)U$ and $\frac{1}{\sqrt{n}}\{Z'(\Sigma^{-1} \otimes P)\epsilon - \mathrm{E}[U'(\Sigma^{-1} \otimes P)\epsilon]\} = \frac{1}{\sqrt{n}}F'(\Sigma^{-1} \otimes I)\epsilon - \frac{1}{\sqrt{n}}F'[\Sigma^{-1} \otimes (I - P)]\epsilon + \frac{1}{\sqrt{n}}\{U'(\Sigma^{-1} \otimes P)\epsilon - \mathrm{E}[U'(\Sigma^{-1} \otimes P)\epsilon]\}$. As $F = \mathrm{diag}\{F_1, F_2\}$ and $U = \mathrm{diag}\{U_1, U_2\}$, where $U_1 = J[S^{-1}u_2, GS^{-1}u_1, GS^{-1}u_2, 0]$, $U_2 = J[S^{-1}u_1, GS^{-1}u_2, GS^{-1}u_1, 0]$, $u_1 = (I - \lambda_{22}G)\epsilon_1 + (\phi_1 I + \lambda_{21}G)\epsilon_2$, and $u_2 = (I - \lambda_{11}G)\epsilon_2 + (\phi_2 I + \lambda_{12}G)\epsilon_1$, it follows by Lemma C.3 and $K/n \to 0$ that $\frac{1}{n}Z'(\Sigma^{-1} \otimes P)Z = \frac{1}{n}F'(\Sigma^{-1} \otimes I)F + O(\max_{i,j}\{\Delta_{ij}\}) + O_p(\sqrt{K/n}) = H + o_p(1)$ and $\frac{1}{\sqrt{n}}\{Z'(\Sigma^{-1} \otimes P)\epsilon - \mathrm{E}[U'(\Sigma^{-1} \otimes P)\epsilon]\} = \frac{1}{\sqrt{n}}F'(\Sigma^{-1} \otimes I)\epsilon + O_p(\max_{i,j}\{\sqrt{\Delta_{ij}}\}) + O_p(\sqrt{K/n}) = \frac{1}{\sqrt{n}}F'(\Sigma^{-1} \otimes I)\epsilon + o_p(1)$. As $\frac{1}{\sqrt{n}}F'(\Sigma^{-1} \otimes I)\epsilon \xrightarrow{d} N(0, H)$ by Theorem A in Kelejian and Prucha (1999), it follows by Slutsky's theorem that $\sqrt{n}(\tilde{\delta}_{3sls} - \delta - b_{3sls}) \xrightarrow{d} N(0, H^{-1})$. As $\frac{1}{n}\mathrm{E}[U'(\Sigma^{-1} \otimes P)\epsilon]$ given by (40) is $O(K/n)$ by Lemma C.2, $b_{3sls} = O_p(K/n)$.

Then, to obtain the asymptotic distribution of the feasible 3SLS $\hat{\delta}_{3sls}$, it is sufficient to show that $\sqrt{n}(\hat{\delta}_{3sls} - \tilde{\delta}_{3sls}) = o_p(1)$. $\sqrt{n}(\hat{\delta}_{3sls} - \tilde{\delta}_{3sls}) = [\frac{1}{n}Z'(\tilde{\Sigma}^{-1} \otimes P)Z]^{-1}\frac{1}{n}Z'[\sqrt{n}(\tilde{\Sigma}^{-1} - \Sigma^{-1}) \otimes P]\epsilon - [\frac{1}{n}Z'(\tilde{\Sigma}^{-1} \otimes P)Z]^{-1}\{\frac{1}{n}Z'[\sqrt{n}(\tilde{\Sigma}^{-1} - \Sigma^{-1}) \otimes P]Z\}[\frac{1}{n}Z'(\Sigma^{-1} \otimes P)Z]^{-1}\frac{1}{n}Z'(\Sigma^{-1} \otimes P)\epsilon$. As $\sqrt{n}(\tilde{\Sigma}^{-1} -$

$\Sigma^{-1}) = O_p(1)$, it follows by a similar argument as above that $\frac{1}{n}Z'(\Sigma^{-1} \otimes P)Z = O_p(1)$, $\frac{1}{n}Z'(\tilde{\Sigma}^{-1} \otimes P)Z = O_p(1)$, and $\frac{1}{n}Z'[\sqrt{n}(\tilde{\Sigma}^{-1} - \Sigma^{-1}) \otimes P]Z = O_p(1)$. On the other hand, $\frac{1}{n}Z'[\sqrt{n}(\tilde{\Sigma}^{-1} - \Sigma^{-1}) \otimes P]\epsilon = \frac{1}{n}F'[\sqrt{n}(\tilde{\Sigma}^{-1} - \Sigma^{-1}) \otimes P]\epsilon + \frac{1}{n}U'[\sqrt{n}(\tilde{\Sigma}^{-1} - \Sigma^{-1}) \otimes P]\epsilon = O_p(\sqrt{K/n}) = o_p(1)$ and $\frac{1}{n}Z'(\Sigma^{-1} \otimes P)\epsilon = \frac{1}{n}F'(\Sigma^{-1} \otimes P)\epsilon + \frac{1}{n}U'(\Sigma^{-1} \otimes P)\epsilon = O_p(\sqrt{K/n}) = o_p(1)$. Therefore, $\sqrt{n}(\hat{\delta}_{3sls} - \tilde{\delta}_{3sls}) = o_p(1)$ and the desired result follows. ■

Proof of Proposition 9. From the proof of Proposition 8, it is sufficient to show that $\sqrt{n}(\hat{b}_{3sls} - b_{3sls}) = R_1 - R_2 = o_p(1)$, where $R_1 = [\frac{1}{n}Z'(\tilde{\Sigma}^{-1} \otimes P)Z]^{-1}\frac{1}{\sqrt{n}}\{\hat{\mathrm{E}}[U'(\Sigma^{-1} \otimes P)\epsilon] - \mathrm{E}[U'(\Sigma^{-1} \otimes P)\epsilon]\}$ and $R_2 = [\frac{1}{n}Z'(\tilde{\Sigma}^{-1} \otimes P)Z]^{-1}\{\frac{1}{n}Z'[\sqrt{n}(\tilde{\Sigma}^{-1} - \Sigma^{-1}) \otimes P]Z\}[\frac{1}{n}Z'(\Sigma^{-1} \otimes P)Z]^{-1}\frac{1}{n}\mathrm{E}[U'(\Sigma^{-1} \otimes P)\epsilon]$. By a similar argument as in the proof of Proposition 7, $\frac{1}{\sqrt{n}}\{\hat{\mathrm{E}}[U'(\Sigma^{-1} \otimes P)\epsilon] - \mathrm{E}[U'(\Sigma^{-1} \otimes P)\epsilon]\} = O_p(K/n)$. By a similar argument as in the proof of Proposition 8, $\frac{1}{n}Z'(\Sigma^{-1} \otimes P)Z = O_p(1)$, $\frac{1}{n}Z'(\tilde{\Sigma}^{-1} \otimes P)Z = O_p(1)$, and $\frac{1}{n}Z'[\sqrt{n}(\tilde{\Sigma}^{-1} - \Sigma^{-1}) \otimes P]Z = O_p(1)$. By Lemma C.2, $\frac{1}{n}\mathrm{E}[U'(\Sigma^{-1} \otimes P)\epsilon] = O(K/n) = o(1)$. Therefore, $R_1 = o_p(1)$ and $R_2 = o_p(1)$ as $K/n \to 0$. ■

References

Ballester, C., Calvó-Armengol, A. and Zenou, Y. (2006). Who's who in networks. wanted: the key player, *Econometrica* **74**: 1403–1417.

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

Blume, L., Brock, W., Durlauf, S. and Ioannides, Y. (2011). Identification of social interactions, *Handbook of Social Economics* **1**: 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. (2012). Multivariate choice and identification of social interactions. CEPR discussion paper No. 9159.

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

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

Hausman, J., Newey, W., Woutersen, T., Chao, J. and Swanson, N. (2008). Instrumental variable estimation with heteroscedasticity and many instruments. Working paper.

Horn, R. and Johnson, C. (1985). *Matrix Analysis*, Cambridge Univsersity Press, Cambridge.

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. (2003). Best spatial two-stage least squares estimators for a spatial autoregressive model with autoregressive disturbances, *Econometric Reviews* **22**: 307–335.

Lee, L. F. (2007). Identification and estimation of econometric models with group interactions, contextual factors and fixed effects, *Journal of Econometrics* **140**: 333–374.

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

Liu, X., Patacchini, E. and Zenou, Y. (2012). Peer effects: Social multiplier or social norms? Working paper, University of Colorado.

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.