Identification and Efficient Estimation of Simultaneous Equations Network Models[*]

Xiaodong Liu[†]

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

July, 2014

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, the variation in the Bonacich centrality across nodes in a network can be used as an IV to identify social interaction effects and improve estimation efficiency. The number of such IVs 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 IVs. We propose a bias-correction procedure for the many-instrument bias. Simulation experiments show that the bias-corrected estimators perform well in finite samples. We also provide an empirical example to illustrate the proposed estimation procedure.

JEL classification: C31, C36

Key words: social networks, simultaneous equations, identification, efficient estimation, many-instrument bias

1 Introduction

Since the seminal work by Manski (1993), social network models have 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 may be influenced by peers' choices (*the endogenous effect*), by peers' 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 choices 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 by 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* choice and *average* characteristics of an agent's connections (or friends).[1] Based on the important observation that in a social network, an agent's friend's friend may not be a (direct) 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) consider the estimation of the *local-aggregate* social network model, where the endogenous effect is given by the *aggregate* choice 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 an additional instrumental variable (IV) to improve estimation efficiency. Liu et al. (2014) give the identification condition for the local-aggregate model and show that the condition is in general weaker than that for the local-average model derived by Bramoullé et al. (2009). They also propose a J test for the specification of network models.

The above mentioned papers focus on single-equation network models involving only one activity. However, in real life, an agent's decision usually involves more than one activity. For example, a

---

[1]Bramoullé et al. (2009) indicate that the first part of Proposition 1 applies to arbitrary (non-row-normalized) matrices.

student may need to balance time between study and extracurriculars and a firm may need to allocate resources between production and R&D (König et al., 2014). 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 choice in a certain activity may depend on his/her choice in a related activity, and *the cross-activity peer effect*, where an agent's choice in a certain activity may depend on peers' choices in a related activity. Cohen-Cole et al. (2012) derive the identification conditions for the various social interaction effects and generalize the spatial two-stage least-squares (2SLS) and three-stage least-squares (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 can be weaker than that for the local-average model. For efficient estimation, we suggest 2SLS and 3SLS estimators using additional IVs based on the Bonacich centrality of each network. As the number of such IVs 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. We also provide an empirical example to illustrate the proposed estimation procedure.

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 provides an empirical example. Section 8 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 $G = [g_{ij}]$. If the network is undirected, then G is symmetric with $g_{ij} = 1$ if i and j are connected and $g_{ij} = 0$ otherwise. If the network is directed, then G may be asymmetric with $g_{ij} = 1$ if an arc is directed from i to j and $g_{ij} = 0$ otherwise.[2] 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 undirected star-shaped network.
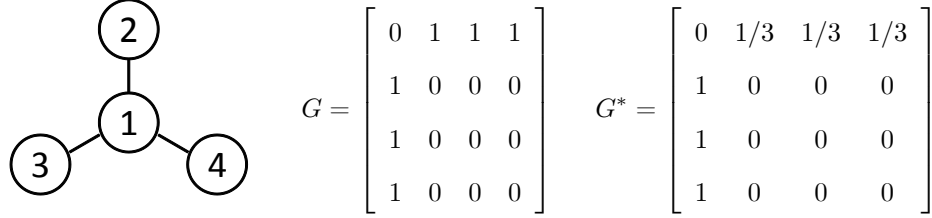

$$G = \begin{bmatrix} 0 & 1 & 1 & 1 \\ 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \end{bmatrix} \quad G^* = \begin{bmatrix} 0 & 1/3 & 1/3 & 1/3 \\ 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \end{bmatrix}$$

Figure 1: an example of G and G^* for a undirected 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

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

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

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

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

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

---

[2] The undirected network can be considered as a special case of the directed network. The identification result and estimation method of the paper hold for both directed and undirected networks.

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

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

which may indicate strategic substitutability or complementarity depending on the signs of λ_{21}^* and λ_{12}^*.

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

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

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

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

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

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

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

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

5

If S is nonsingular,[3] then the local-aggregate network game has a unique Nash equilibrium in pure strategies with the equilibrium efforts given by

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

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

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

$$u(y_{1i}, y_{2i}) = \pi_{1i}^* y_{1i} + \pi_{2i}^* y_{2i} - \frac{1}{2}\phi_1^* y_{1i}^2 - \frac{1}{2}\phi_2^* y_{2i}^2 + \phi^* y_{1i} y_{2i} \tag{9}$$

$$+ \lambda_{11}^* \sum_{j=1}^n g_{ij}^* y_{1i} y_{1j} + \lambda_{22}^* \sum_{j=1}^n g_{ij}^* y_{2i} y_{2j} + \lambda_{21}^* \sum_{j=1}^n g_{ij}^* y_{1i} y_{2j} + \lambda_{12}^* \sum_{j=1}^n g_{ij}^* y_{2i} y_{1j}.$$

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

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

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

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

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

---

[3]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.

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. (2014), 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.[4] 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

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

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

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

$$Y_1^* = S^{-1}[(I - \lambda_{22}G)\pi_1 + (\phi_1 I + \lambda_{21}G)\pi_2]l_n,$$

$$Y_2^* = S^{-1}[(I - \lambda_{11}G)\pi_2 + (\phi_2 I + \lambda_{12}G)\pi_1]l_n.$$

Thus, the agents would have different equilibrium efforts and payoffs if Gl_n is not proportional to l_n.[5]

---

[4]Liu et al. (2014) 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.

[5]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

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

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

Therefore, the local-aggregate and local-average network games have different equilibrium and policy implications. Understanding the different micro-foundations of the two models, including their different equilibrium implications, may help us to understand which model is more appropriate for a certain economic context.[6] 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 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 r-th network. For the r-th network, let

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

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.[7] It follows by (4) and (5) that $Y_{1,r}$ and $Y_{2,r}$, which are $n_r \times 1$ vectors of observed choices of two related activities for the agents in the r-th network, are given by

$$
\begin{aligned}
Y_{1,r} &= \phi_1 Y_{2,r} + \lambda_{11} G_r Y_{1,r} + \lambda_{21} G_r Y_{2,r} + X_r \beta_1 + G_r X_r \gamma_1 + \alpha_{1,r} l_{n_r} + \epsilon_{1,r}, \\
Y_{2,r} &= \phi_2 Y_{1,r} + \lambda_{22} G_r Y_{2,r} + \lambda_{12} G_r Y_{1,r} + X_r \beta_2 + G_r X_r \gamma_2 + \alpha_{2,r} l_{n_r} + \epsilon_{2,r}.
\end{aligned}
$$

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 $Y_k = (Y'_{k,1}, \cdots, Y'_{k,\bar{r}})'$, $X = (X'_1, \cdots, X'_{\bar{r}})'$, $\alpha_k = (\alpha_{k,1}, \cdots, \alpha_{k,\bar{r}})'$,

---

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

[6] It may not always obvious to see whether the local-average or local-aggregate model is more appropriate just by economic intuition. Sometimes, a specification test is needed. Liu et al. (2014) extend the J test to network models.

[7] The specification of $\Pi_{1,r}$ and $\Pi_{2,r}$ is quite common for network models (see, e.g., Calvó-Armengol et al., 2009; Bramoullé et al., 2009; Liu and Lee, 2010). In spatial econometrics literature, this specification is known as the spatial Durbin model (LeSage and Pace, 2009). Essentially this model specification imposes exclusion restrictions such that $\Pi_{1,r}$ and $\Pi_{2,r}$ do not depend on the characteristics of indirect connections.

$\epsilon_k = (\epsilon'_{k,1}, \cdots, \epsilon'_{k,\bar{r}})'$, $L = \text{diag}\{l_{n_r}\}_{r=1}^{\bar{r}}$ and $G = \text{diag}\{G_r\}_{r=1}^{\bar{r}}$. Then, for all the $\bar{r}$ networks,

$$Y_1 = \phi_1 Y_2 + \lambda_{11} GY_1 + \lambda_{21} GY_2 + X\beta_1 + GX\gamma_1 + L\alpha_1 + \epsilon_1, \tag{14}$$

$$Y_2 = \phi_2 Y_1 + \lambda_{22} GY_2 + \lambda_{12} GY_1 + X\beta_2 + GX\gamma_2 + L\alpha_2 + \epsilon_2. \tag{15}$$

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

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

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

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.

3.2.1 Endogenous effect and contextual effect

The endogenous effect, where an agent's choice may depend on choices of his/her friends in the same activity, is captured by the coefficients λ_{11} and λ_{22}. The contextual effect, where an agent's choice 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 directly 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 IVs to identify the endogenous effect from the contextual effect in the

local-average model.[8]

On the other hand, when G_r does not have constant row sums, the number of friends represented by the element of $G_r l_{n_r}$ 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 IV 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.

3.2.2 Simultaneity effect and cross-activity peer effect

The simultaneity effect, where an agent's choice in an activity may depend on his/her choice in a related activity, can be seen in the coefficients ϕ_1 and ϕ_2. The cross-activity peer effect, where an agent's choice may depend on those of his/her friends in a related activity, is represented by the coefficients λ_{21} and λ_{12}.

For a standard simultaneous equations model without social interaction effects, simultaneity leads to a well known identification problem 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 $J, JG^*, JG^{*2}, JG^{*3}$ 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 (but not both) can be identified under weaker conditions.

However, neither the intransitivity in G nor the variation in Bonacich centrality would 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.

3.2.3 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*

---

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

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 of the same agent in 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 IVs.

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 5) 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. For ease of presentation, we assume G and X are nonstochastic. Furthermore, in this section, we assume X is a column vector.

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

$$Y_1 = \lambda_{11}GY_1 + X\beta_1 + GX\gamma_1 + L\alpha_1 + \epsilon_1, \tag{22}$$

$$Y_2 = \lambda_{22}GY_2 + X\beta_2 + GX\gamma_2 + L\alpha_2 + \epsilon_2. \tag{23}$$

For the SUR network model, an agent's choice is still allowed to be correlated with his/her own choices in 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

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

For (22), let $Z_1 = [GY_1, X, GX]$. Identification of (22) requires $E(JZ_1) = [E(JGY_1), JX, JGX]$ to have full column rank. As $(I - \lambda_{11}G)^{-1} = I + \lambda_{11}G(I - \lambda_{11}G)^{-1}$ and $(I - \lambda_{11}G)^{-1}G = G(I - \lambda_{11}G)^{-1}$, it follows from (24) that

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

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 IVs 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 IVs for identification. The following proposition gives a sufficient condition for $E(JZ_1)$ to have full column rank.

Proposition 1 *Suppose G_r has non-constant row sums for some network r. For equation (22), $E(JZ_1)$ 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) *$G_r^2 = \rho_1 I_{n_r} + \rho_2 G_r$ and $\alpha_{1,r}(1 - \rho_2\lambda_{11} - \rho_1\lambda_{11}^2) \neq 0$.*

The identification condition for the *local-aggregate* SUR model given in Proposition 1 is, in general, weaker than that for the *local-average* SUR model. As pointed out by Bramoullé et al.

(2009),[9] 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, $G = \text{diag}\{G_r\}_{r=1}^{\bar{r}}$ 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, by Proposition 1(i), the local-aggregate SUR model can be identified if $\alpha_{1,r} \neq 0$ or $\lambda_{11}\beta_1 + \gamma_1 \neq 0$.

$$
\text{1} \longleftarrow \text{2} \longrightarrow \text{3} \qquad G_r = \begin{bmatrix} 0 & 0 & 0 \\ 1 & 0 & 1 \\ 0 & 0 & 0 \end{bmatrix}
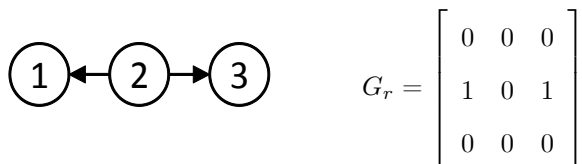$$

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

Figure 2 gives an example where the condition in Proposition 1(ii) is satisfied. For the directed graph in Figure 2, 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, by Proposition 1(ii), the local-aggregate SUR model is identified if $\alpha_{1,r} \neq 0$.

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 is allowed to depend on his/her own choice in a related activity. However, we assume that an agent's choice is not affected by the friends' choices in related activities. Under the restrictions $\lambda_{21} = \lambda_{12} = 0$, (14) and (15)

---

[9]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.

become

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

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

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 are weaker in general. 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 $G^* = \mathrm{diag}\{G_r^*\}_{r=1}^{\bar{r}}$, 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 r-th 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 according to Proposition 2(iii).

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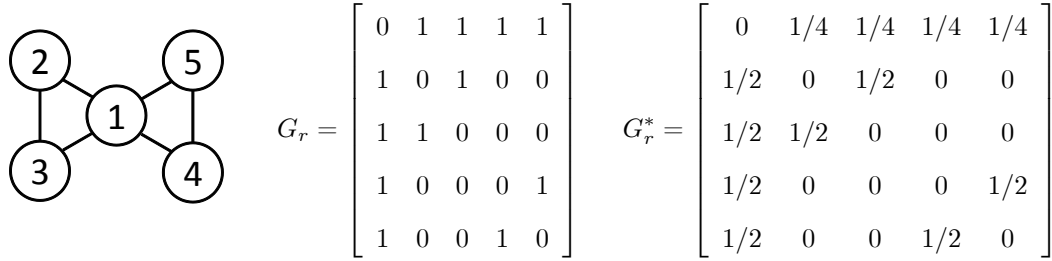$$

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 according to Proposition 2(iv).

Figure 4 provides an example where the condition in Proposition 2(ii) is 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 according to Proposition 2(ii).

$$
G_r = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{bmatrix}
$$
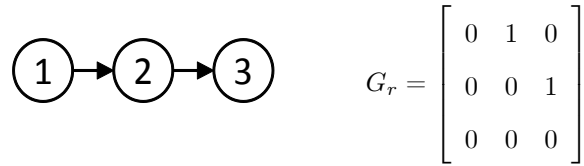
Figure 4: an example where the local-aggregate model can be identified by Proposition 2(ii).

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 is allowed to depend on his/her friends' choices 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

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

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

For (27), let $Z_1 = [GY_1, GY_2, X, GX]$. The following proposition gives sufficient conditions for $\text{E}(JZ_1)$ 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, $\text{E}(JZ_1)$ 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 $G_r^2 l_{n_r} = \eta_1 l_{n_r} + \eta_2 G_r l_{n_r}$, $\text{E}(JZ_1)$ 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 conditions for the *local-aggregate* model given by Proposition 3 are weaker in general. 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 according to 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 according to 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 $E(\bar{Z}_1)$ and $E(\bar{Z}_2)$, 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), $E(\bar{Z}_1)$ and $E(\bar{Z}_2)$ 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 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 column vectors and $[X_1, X_2]$ has full column rank.[10] From the reduced-form equations (16) and (17), we have

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

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

where S is 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 $E(JZ_1)$ 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, $E(JZ_1)$ 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.*

---

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

- 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 $D_3^{\ddagger}$ 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 according to Proposition 5(iv).

5   Estimation

5.1   The 2SLS estimator with many IVs

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 IVs 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 IVs 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 3), 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 IVs F_1 and F_2 can be approximated by a linear combination of an $n \times K$ 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 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 $\hat{\delta}_{2,2sls} = (Z_2'P_KZ_2)^{-1}Z_2'P_KY_2$.[11]

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 IVs 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 (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 $\mathrm{E}[\frac{1}{n}Z_2'P_K(Y_2-$

---

[11] The finite sample properties of IV-based estimators can be sensitive to the number of IVs. In a recent paper, Liu and Lee (2013) derive the Nagar-type approximate MSE of the 2SLS estimator for spatial models, which can be minimized to choose the optimal number of IVs as in Donald and Newey (2001). The approximate MSE can be derived in a similar way for the proposed estimators in the paper.

$Z_2\delta_2)] = \frac{1}{n}\mathrm{E}(U_2'P_K\epsilon_2)$, where

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

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

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 IVs 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 $\bar{r}/n = 1/\bar{n}_r \to 0$, where $\bar{n}_r = n/\bar{r}$ 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 $\bar{r}^2/n = \bar{r}/\bar{n}_r \to 0$. 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 IVs (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 $\hat{b}_{1,2sls} = (Z_1'P_KZ_1)^{-1}\hat{\mathrm{E}}(U_1'P_K\epsilon_1)$ and $\hat{b}_{2,2sls} = (Z_2'P_KZ_2)^{-1}\hat{\mathrm{E}}(U_2'P_K\epsilon_2)$, 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

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

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 $\sqrt{n}(\hat{\delta}_{2,bc2sls} - \delta_2) \xrightarrow{d} N(0, \sigma_2^2 H_{22}^{-1})$.*

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

5.2 The 3SLS estimator with many IVs

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 extend the 3SLS estimator proposed by Kelejian and Prucha (2004) to estimate the local-aggregate network model in the presence of many IVs.

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 given by (38). As $\text{E}(\epsilon\epsilon') = \Sigma \otimes I$, 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 IVs goes to infinity at a rate slower than the sample size, the 3SLS estimator is consistent and asymptotically normal with an asymptotic bias given by $\sqrt{n}b_{3sls}$. 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 IVs 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

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

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

For the experiment, we consider four samples with different numbers of networks $\bar{r}$ and network sizes n_r. Samples 1, 2 and 3 have $\bar{r} = 30$ equal-sized networks with $n_r = 10$, $n_r = 15$ and $n_r = 30$ respectively. Sample 4 has $\bar{r} = 60$ equal-sized networks with $n_r = 15$. For each network, the adjacency matrix G_r is generated in a similar way as in Liu and Lee (2010). First, for the i-th row of G_r ($i = 1, \cdots, n_r$), we generate an integer p_{ri} uniformly at random from the set of integers $\{1, 2, 3\}$. Then, if $i + p_{ri} \leq n_r$, we set the $(i + 1)$-th, $\cdots$, $(i + p_{ri})$-th entries of the i-th 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 $(i + p_{ri} - n_r)$ entries of the i-th row will be ones. We choose this design to generate G_r as it allows us to easily control the parameter space and the variance of row sums.[12]

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 $\phi_1 = \phi_2 = 0.2$ and $\lambda_{11} = \lambda_{22} = \lambda_{12} = \lambda_{21} = 0.1$.[13] We consider different values for β and γ in the experiment.

We consider the following estimators in the simulation experiment: (i) 2SLS-1 and 3SLS-1 with the IV matrix $Q_1 = J[X_1, GX_1, G^2 X_1, X_2, GX_2, G^2 X_2]$; (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 numbers of friends given by GL as additional IVs to improve estimation efficiency. Note that GL has $\bar{r}$ columns. So the number

---

[12] Note that the parameter space of ϕ's and λ's depends on $||G||_\infty$ (see footnote 3). 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.

[13] We choose ϕ's and λ's so that the S is invertible according to footnote 3.

of IVs in Q_2 increases with the number of networks.

The estimation results of equation (41) are reported in Tables 1-8. We report the mean and standard deviation (SD) of the empirical distributions of the estimates. To facilitate the comparison of different estimators, we also report their root mean square errors (RMSE). The main observations from the experiment are summarized as follows.

[Insert Tables 1-8 here]

(a) The additional IVs based on the numbers 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 1-4) and the correlation across equations is weak ($\sigma_{12} = 0.1$), for the sample with $n_r = 10$ 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 3.9%, 17.1%, 14.7%, 1.5%, and 5.4%. 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 13.7%, 22.9%, 20.6%, 10.6%, and 10.9%. 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 5-8). When $\sigma_{12} = 0.1$ (see the top panel of Table 5), SD reductions of 2SLS-2 estimators of $\phi_1, \lambda_{11}, \lambda_{21}, \beta_1, \gamma_1$ are, respectively, about 18.1%, 37.5%, 33.8%, 6.0%, and 12.3%. The SD reduction of the 3SLS estimator with Q_2 follows a similar pattern.

(b) The additional IVs 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 1, bias reductions of BC3SLS estimators of $\phi_1, \lambda_{11}, \lambda_{21}$ (relative to 3SLS-2) are, respectively, 100.0%, 100.0%, and 66.7%, when $\sigma_{12} = 0.1$.

(d) The 3SLS estimator improves the efficiency upon the 2SLS estimator. The improvement is more 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 8.5%, 9.7%, 3.2%, 25.4%, and 14.5%.

7 Empirical Example

To illustrate the proposed estimators, we study the peer effect of sporting activities on juvenile delinquency using a unique and now widely used data set provided by the National Longitudinal Study of Adolescent Health (Add Health). The Add Health data provides national representative information on 7th-12th graders in the United States. The in-school survey was conducted during the 1994-1995 year with four follow-up in-home interviews. Here we only use the first wave of Add Health data.

In this empirical example, we consider the estimation of (29) and (30) where Y_1 and Y_2 are indices of participation in delinquent conducts and sporting activities respectively. To be more specific, the Add Health survey reports the frequency of the following delinquent behaviors in the past 12 months, namely, smoking cigarettes; drinking beer, wine, or liquor; getting drunk; racing on a bike, on a skateboard or roller blades, or in a boat or car; doing something dangerous due to dare; lying to parents or guardians; and skipping school without an excuse, coded using an ordinal scale as 0 (never), 1 (once or twice), 2 (once a month or less), 3 (2 or 3 days a month), 4 (once or twice a week), 5 (3-5 days a week), and 6 (nearly everyday). Y_1 is given by the average frequency of the above delinquent behaviors.[14] Y_2 corresponds to the average frequency of participating the following sporting activities in the past week, including going roller-blading, roller-skating, skate-boarding, or bicycling; playing an active sport, such as baseball, softball, basketball, soccer, swimming, or football; and doing exercise, such as jogging, walking, karate, jumping rope, gymnastics or dancing, coded as 0 (not at all), 1 (1 or 2 times), 2 (3 or 4 times), and 3 (5 or more times).

The adjacency matrix $G = [g_{ij}]$ is constructed based on the friend-nomination information in the Add Health data. In the survey, the respondent is asked to nominate his/her best friends (up to 5 male friends and 5 female friends) from a school roster.[15] We assume friendship is reciprocal. Thus, for students i and j, $g_{ij} = 1$ if either i nominates j or j nominates i as a friend and $g_{ij} = 0$ otherwise.[16] After removing isolated students and pairs (i.e. network with only two students), the sample consists of 8,236 students distributed over 515 networks[17], with network size ranging

---

[14]In the sample considered, less than 5% of the students claimed they never committed the above delinquent conducts in the past 12 months.

[15]Less than 1% of the students in the sample nominated 10 friends and thus the bound on the number of friend-nominations is not binding.

[16]We also consider an alternative specification of the adjacency matrix $G = [g_{ij}]$ without assuming friendship is receprocal so that $g_{ij} = 1$ if i nominates j as a friend and $g_{ij} = 0$ otherwise. The estimation result is similar to the one reported here.

[17]Here, a network is defined as the largest set of students who are directly or indirectly connected through friend-

from 3 to 949. Because the strength of peer effect may vary with network sizes (Calvó-Armengol et al., 2009), we exclude networks larger than the 99th percentile of the network size distribution. Our selected sample consists of 5,947 students distributed over 509 networks, with network sizes ranging from 3 to 140. The mean and the standard deviation of network size are 11.68 and 22.62. Furthermore, in our sample, the average number of friends of a student is 2.74 with the standard deviation 2.09. A summary of the data can be found in Table 9.

[Insert Table 9 here]

From the discussion in Subsection 4.2.3, we understand to identify the peer effects in (29) and (30), we need to introduce some IVs (or exclusion restrictions), at least one for each activity. The IVs for the delinquency index Y_1 of a student are indicator variables of whether the parent smokes cigarettes and whether the parent took the student to a religious service in the past four weeks. The IVs for participation in sporting activities Y_2 are indicator variables of whether whether the parent played a sport with the student in the past four weeks and whether the student is well coordinated. The coefficient estimates of the IVs are all statistically significant with signs consistent with our expectations. The first-stage F test statistics also suggest the IVs are relevant and informative. The p-values of over-identification restrictions (OIR) test are larger than the conventional significance levels, lending some evidence to the validity of the IVs.

[Insert Table 10 here]

We consider the 2SLS and 3SLS estimators with the IV matrix $Q_1 = J[X, GX, G^2X]$, where X contains the linearly independent columns in $[X_1, X_2]$, and the BC2SLS and BC3SLS estimators with the IV matrix $Q_2 = [Q_1, JGL]$, where JGL are additional IVs based on the numbers of friends. The estimation result of equation (29) is reported in Table 10.[18] The estimates from different estimators are largely consistent with each other. Interestingly, we can see that delinquent behavior of a student increases delinquent behavior of his/her friend ($\hat{\lambda}_{11} > 0$). On the other hand, participation in sporting activities not only reduces a student's own delinquent behavior ($\hat{\phi}_1 < 0$) but also reduces delinquent behavior of his/her friend ($\hat{\lambda}_{12} < 0$). Finally, we conduct a J test as in Liu et al. (2014)

---

nomination. By this definition, students from two different networks cannot be friends.

[18]To save space, we only report the first-stage F test statistic from the estimation of equation (30). The estimation result of equation (30) is available upon request.

against the alternative specification of a local-average model. The p-value of the J test is 0.823, suggesting that the local-aggregate model is indeed a better model for the data.

8 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 in general weaker than those for the model with a row-normalized adjacency matrix.

For efficient estimation, we propose 2SLS and 3SLS estimators for the simultaneous equations network model using a set of feasible IVs to approximate the best (infeasible) IVs given by the reduced-form equations of the model. When the network adjacency matrix is not row-normalized, the set of feasible IVs includes the leading order terms of the Bonacich centrality for each network, and thus the number of IVs depends on the number of networks. When there are many networks in the data, we would have many IVs. We show that the proposed 2SLS and 3SLS estimators are consistent and asymptotically normally distributed (with an asymptotic bias) when the number of IVs 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 IVs grows slower than the sample size (or, when the average network size is large). Monte Carlo experiments show that the IVs 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.

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.[19]

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 3 requires the row and column sums of the adjacency matrix G to be uniformly bounded, which in turn requires the the number of *direct* links of each node in a network to be bounded for the *local-aggregate* model. Intuitively, this assumption says, as the sample size increases, the number of "close" friends of a person cannot go to infinity. Note that, Assumption 3 does not require the network size to be bounded. In fact, the network size is allowed to go to infinity as the sample size increases. In another word, the number of "indirect" friends of a person (i.e. friends'

---

[19]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)

friends, friends' friends' friends, etc.) is allowed to go to infinity. From this perspective, Assumption 3 is not too restrictive.

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 IVs increases with the sample size. This condition is commonly assumed in the many-instrument literature (see, eg., Donald and Newey, 2001; Hansen et al., 2008; Hausman et al., 2012).

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,

$$
\begin{align}
D_2 &= [A_2', C_2']' \tag{47} \\
D_2^* &= [A_2^{*\prime}, C_2']' \tag{48} \\
D_2^\dagger &= [A_2', C_2^{*\prime}]' \tag{49} \\
D_2^\ddagger &= [A_2^{*\prime}, C_2^{*\prime}]'. \tag{50}
\end{align}
$$

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,

$$D_3 = [A_3', B_3', C_3']' \tag{51}$$

$$D_3^* = [A_3^{*'}, B_3^{*'}, C_3']' \tag{52}$$

$$D_3^\dagger = [A_3', B_3', C_3^{*'}]' \tag{53}$$

$$D_3^\ddagger = [A_3^{*'}, B_3^{*'}, C_3^{*'}]' \tag{54}$$

$$D_3^\sharp = [A_3^{**'}, B_3^{**'}, C_3^{*'}]'. \tag{55}$$

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 $\|A\| = \sqrt{\text{tr}(A'A)}$ 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) $\text{tr}(P) = K$. (ii) *Suppose that $\{A\}$ is a sequence of $n \times n$ UB matrices. For $B = PA$, $\text{tr}(B) = O(K)$, $\text{tr}(B^2) = O(K)$, 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) $\Delta_{ij} = \frac{1}{n}\text{tr}(F_i'(I - P)F_j) = o(1)$; (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) $\frac{1}{\sqrt{n}}[\epsilon_i'PA\epsilon_j - \text{E}(\epsilon_i'PA\epsilon_j)] = O_p(\sqrt{K/n})$.*

Proof. For (i), $\frac{1}{n}\text{tr}(F_i'(I-P)F_j) = \frac{1}{n}\text{tr}((F_i - Q\pi_i)'(I-P)(F_j - Q\pi_j)) \le [\frac{1}{n}\text{tr}((F_i - Q\pi_i)'(F_i - Q\pi_i))]^{1/2}[\frac{1}{n}\text{tr}((F_j - Q\pi_j)'(I-P)(F_j - Q\pi_j))]^{1/2} \le [\frac{1}{n}\text{tr}((F_i - Q\pi_i)'(F_i - Q\pi_i))]^{1/2}[\frac{1}{n}\text{tr}((F_j - Q\pi_j)'(F_j - Q\pi_j))]^{1/2} = [\frac{1}{n}||F_i - Q\pi_i||^2]^{1/2}[\frac{1}{n}||F_j - Q\pi_j||^2]^{1/2} \to 0$, by Assumption 5 and Lemma C.1. For (ii) and (iii), see Liu and Lee (2010). For (iv), as $\text{Var}[\frac{1}{\sqrt{n}}F_i'(I-P)A\epsilon_j] = \frac{\sigma_j^2}{n}F_i'(I-P)AA(I-P)F_i \le \frac{\sigma_j^2}{n}F_i'(I-P)F_i||AA||_\infty = O(\Delta_{ii})$, 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 $M = PA$. $\text{E}(\epsilon_j'M'\epsilon_i\epsilon_i'M\epsilon_j) - \text{E}(\epsilon_j'M'\epsilon_i)\text{E}(\epsilon_i'M\epsilon_j) = c_1 \sum_s M_{ss}^2 + c_2\text{tr}(M^2) + c_3\text{tr}(M'M)$, 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 $\sum_s M_{ss}^2, \text{tr}(M^2), \text{tr}(M'M)$ are $O(K)$ by Lemma C.2, we have $\frac{1}{\sqrt{n}}[\epsilon_i'PA\epsilon_j - \text{E}(\epsilon_i'PA\epsilon_j)] = O_p(\sqrt{K/n})$ by Markov's inequality. ∎

D Proofs

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

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

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

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

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

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

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 $\alpha_{1,r} \ne 0$ or $\lambda_{11}\beta_1 + \gamma_1 \ne 0$. (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 $\alpha_{1,r} \ne 0$ and $1 - \rho_2\lambda_{11} - \rho_1\lambda_{11}^2 \ne 0$. ∎

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 i-th 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 $G^2 = \rho_1 I + \rho_2 G$, 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,

$$J_r[\mathrm{E}(Y_{2,r})d_1 + \mathrm{E}(G_r Y_{1,r})d_2 + X_r d_3 + G_r X_r d_4] = 0 \tag{58}$$

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

$$\mathrm{E}(Y_{2,r})d_1 + \mathrm{E}(G_r Y_{1,r})d_2 + X_r d_3 + G_r X_r d_4 + l_{n_r}\mu = 0, \tag{59}$$

where $\mu = -\frac{1}{n_r} l'_{n_r}[\mathrm{E}(Y_{2,r})d_1 + \mathrm{E}(G_r Y_{1,r})d_2 + X_r d_3 + G_r X_r d_4]$. Under the exclusion restrictions $\lambda_{21} = \lambda_{12} = 0$, the reduced-form equations (16) and (17) become

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

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

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 $G_r S_{\lambda,r} = S_{\lambda,r}G_r$, it follows from (60) and (61) that

$$X_r a_1 + G_r X_r a_2 + G_r^2 X_r a_3 + G_r^3 X_r a_4 + l_{n_r}c_1 + G_r l_{n_r}c_2 + G_r^2 l_{n_r}c_3 = 0,$$

where

$$a_1 = (\phi_2\beta_1 + \beta_2)d_1 + (1 - \phi_1\phi_2)d_3$$

$$a_2 = (\phi_2\gamma_1 + \gamma_2 - \lambda_{11}\beta_2)d_1 + (\beta_1 + \phi_1\beta_2)d_2 - (\lambda_{11} + \lambda_{22})d_3 + (1 - \phi_1\phi_2)d_4$$

$$a_3 = -\lambda_{11}\gamma_2 d_1 + (\gamma_1 + \phi_1\gamma_2 - \lambda_{22}\beta_1)d_2 + \lambda_{11}\lambda_{22}d_3 - (\lambda_{11} + \lambda_{22})d_4$$

$$a_4 = -\lambda_{22}\gamma_1 d_2 + \lambda_{11}\lambda_{22}d_4$$

and

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

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

$$c_3 = -\lambda_{22}\alpha_{1,r}d_2 + \lambda_{11}\lambda_{22}\mu.$$

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. $\blacksquare$

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,

$$J_r[\mathrm{E}(G_r Y_{1,r})d_1 + \mathrm{E}(G_r Y_{2,r})d_2 + X_r d_3 + G_r X_r d_4] = 0 \tag{62}$$

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

$$\mathrm{E}(G_r Y_{1,r})d_1 + \mathrm{E}(G_r Y_{2,r})d_2 + X_r d_3 + G_r X_r d_4 + l_{n_r}\mu = 0, \tag{63}$$

35

where $\mu = -\frac{1}{n_r} l'_{n_r}[\text{E}(G_r Y_{1,r})d_1 + \text{E}(G_r Y_{2,r})d_2 + X_r d_3 + G_r X_r d_4]$. Under the exclusion restrictions $\phi_1 = \phi_2 = 0$, the reduced-form equations (16) and (17) become

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

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

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 $G_r S_{\phi,r} = S_{\phi,r}G_r$, it follows from (64) and (65) that

$$
X_r a_1 + G_r X_r a_2 + G_r^2 X_r a_3 + G_r^3 X_r a_4 + l_{n_r}c_1 + G_r l_{n_r}c_2 + G_r^2 l_{n_r}c_3 = 0,
$$

where

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

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 $[l_{n_r}, G_r l_{n_r}, G_r^2 l_{n_r}]$ 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_2 given by (47) 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_2^* given by (48) 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 $D_2^\dagger$ given by (49) 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 $D_2^\ddagger$ given by (50) has full rank. $\blacksquare$

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

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

where

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

$$b_1 = \beta_2 d_1$$

$$b_2 = (\gamma_2 - \lambda_{11}\beta_2)d_1 + \phi_1\beta_2 d_2 + \beta_2 d_3$$

$$b_3 = -\lambda_{11}\gamma_2 d_1 + (\lambda_{21}\beta_2 + \phi_1\gamma_2)d_2 + (\gamma_2 - \lambda_{11}\beta_2)d_3$$

$$b_4 = \lambda_{21}\gamma_2 d_2 - \lambda_{11}\gamma_2 d_3$$

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, $b_{1,2sls} = O_p(K/n)$. $\blacksquare$

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. $\frac{1}{\sqrt{n}}[(\tilde{\sigma}_{12} + \check{\phi}_2 \tilde{\sigma}_1^2)\mathrm{tr}(P\tilde{S}^{-1}) - (\sigma_{12} + \phi_2\sigma_1^2)\mathrm{tr}(PS^{-1})] + \frac{1}{\sqrt{n}}[(\tilde{\lambda}_{12}\tilde{\sigma}_1^2 - \tilde{\lambda}_{11}\tilde{\sigma}_{12})\mathrm{tr}(P\tilde{S}^{-1}G) - (\lambda_{12}\sigma_1^2 - \lambda_{11}\sigma_{12})\mathrm{tr}(PS^{-1}G)]$, is $o_p(1)$, where $\tilde{S} = (1 - \check{\phi}_1\check{\phi}_2)I - (\tilde{\lambda}_{11} + \tilde{\lambda}_{22} + \check{\phi}_1\tilde{\lambda}_{12} + \check{\phi}_2\tilde{\lambda}_{21})G + (\tilde{\lambda}_{11}\tilde{\lambda}_{22} - \tilde{\lambda}_{12}\tilde{\lambda}_{21})G^2$. 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 $\sqrt{n}(\tilde{S}^{-1} - S^{-1}) = \sqrt{n}S^{-1}(S - \tilde{S})\tilde{S}^{-1} = \sqrt{n}(\check{\phi}_1\check{\phi}_2 - \phi_1\phi_2)S^{-1}\tilde{S}^{-1} + [\sqrt{n}(\tilde{\lambda}_{11} - \lambda_{11}) + \sqrt{n}(\tilde{\lambda}_{22} - \lambda_{22}) + \sqrt{n}(\check{\phi}_1\tilde{\lambda}_{12} - \phi_1\lambda_{12}) + \sqrt{n}(\check{\phi}_2\tilde{\lambda}_{21} - \phi_2\lambda_{21})]S^{-1}G\tilde{S}^{-1} + [\sqrt{n}(\tilde{\lambda}_{12}\tilde{\lambda}_{21} - \lambda_{12}\lambda_{21}) - \sqrt{n}(\tilde{\lambda}_{11}\tilde{\lambda}_{22} - \lambda_{11}\lambda_{22})]S^{-1}G^2\tilde{S}^{-1}$, it follows that $\frac{1}{\sqrt{n}}[(\tilde{\sigma}_{12} + \check{\phi}_2\tilde{\sigma}_1^2)\mathrm{tr}(P\tilde{S}^{-1}) - (\sigma_{12} + \phi_2\sigma_1^2)\mathrm{tr}(PS^{-1})] = [\sqrt{n}(\tilde{\sigma}_{12} - \sigma_{12}) + \sqrt{n}(\check{\phi}_2\tilde{\sigma}_1^2 - \phi_2\sigma_1^2)]\frac{1}{n}\mathrm{tr}(P\tilde{S}^{-1}) + (\sigma_{12} + \phi_2\sigma_1^2)\frac{1}{n}\mathrm{tr}(P\sqrt{n}(\tilde{S}^{-1} - S^{-1})) = O_p(K/n)$ 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 $K/n \to 0$. $\blacksquare$

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.

Calvó-Armengol, A., Patacchini, E. and Zenou, Y. (2009). Peer effects and social networks in education, *The Review of Economic Studies* **76**: 1239–1267.

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. (2012). Instrumental variable estimation with heteroscedasticity and many instruments, *Quantitative Economics* **3**: 211–255.

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.

König, M., Liu, X. and Zenou, Y. (2014). R&D neworks: Theory, empirics and policy implications. CEPR discussion paper no. 9872.

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.

LeSage, J. and Pace, R. (2009). *Introduction to Spatial Econometrics*, CRC Press, Boca Raton, FL.

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

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

Liu, X., Patacchini, E. and Zenou, Y. (2014). Endogenous peer effects: Local aggregate or local average? Working paper, University of Colorado Boulder.

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.

Table 1: 2SLS and 3SLS Estimation ($n_r = 10$, $\bar{r} = 30$)

	$\varphi_1 = 0.2$	$\lambda_{11} = 0.1$	$\lambda_{21} = 0.1$	$\beta_1 = 0.8$	$\gamma_1 = 0.8$
$\sigma_{12} = 0.1$					
2SLS-1	.202(.051)[.051]	.100(.035)[.035]	.100(.034)[.034]	.806(.066)[.067]	.803(.056)[.056]
2SLS-2	.223(.049)[.054]	.093(.029)[.030]	.095(.029)[.029]	.799(.065)[.065]	.802(.053)[.053]
BC2SLS	.198(.055)[.055]	.100(.030)[.030]	.101(.030)[.030]	.806(.066)[.066]	.803(.054)[.054]
3SLS-1	.202(.051)[.051]	.100(.035)[.035]	.100(.034)[.034]	.806(.066)[.066]	.804(.056)[.056]
3SLS-2	.216(.053)[.056]	.093(.029)[.030]	.097(.030)[.030]	.800(.065)[.065]	.805(.053)[.053]
BC3SLS	.200(.054)[.054]	.100(.030)[.030]	.101(.030)[.030]	.806(.066)[.066]	.803(.054)[.054]
$\sigma_{12} = 0.5$					
2SLS-1	.203(.051)[.051]	.100(.035)[.035]	.099(.034)[.034]	.806(.066)[.066]	.803(.056)[.056]
2SLS-2	.253(.047)[.071]	.094(.028)[.029]	.086(.028)[.032]	.794(.063)[.063]	.794(.051)[.052]
BC2SLS	.193(.057)[.057]	.101(.030)[.030]	.102(.031)[.031]	.807(.066)[.066]	.804(.054)[.054]
3SLS-1	.203(.051)[.051]	.099(.034)[.034]	.099(.034)[.034]	.806(.062)[.062]	.805(.053)[.053]
3SLS-2	.217(.052)[.055]	.093(.029)[.030]	.095(.029)[.029]	.806(.060)[.060]	.809(.049)[.050]
BC3SLS	.200(.054)[.054]	.100(.030)[.030]	.101(.030)[.030]	.805(.061)[.061]	.803(.051)[.051]
$\sigma_{12} = 0.9$					
2SLS-1	.204(.051)[.051]	.100(.035)[.035]	.099(.034)[.034]	.806(.066)[.066]	.803(.055)[.056]
2SLS-2	.282(.044)[.093]	.095(.027)[.027]	.076(.027)[.036]	.790(.059)[.060]	.785(.049)[.051]
BC2SLS	.188(.059)[.060]	.101(.031)[.031]	.104(.031)[.031]	.808(.067)[.067]	.805(.055)[.055]
3SLS-1	.204(.051)[.051]	.098(.031)[.031]	.099(.033)[.033]	.806(.049)[.050]	.807(.045)[.046]
3SLS-2	.219(.052)[.055]	.092(.027)[.028]	.093(.028)[.029]	.814(.049)[.051]	.815(.046)[.048]
BC3SLS	.202(.054)[.054]	.100(.028)[.028]	.100(.030)[.030]	.805(.050)[.050]	.805(.047)[.047]

Mean(SD)[RMSE]

Table 2: 2SLS and 3SLS Estimation ($n_r = 15$, $\bar{r} = 30$)

	$\varphi_1 = 0.2$	$\lambda_{11} = 0.1$	$\lambda_{21} = 0.1$	$\beta_1 = 0.8$	$\gamma_1 = 0.8$
$\sigma_{12} = 0.1$					
2SLS-1	.202(.039)[.039]	.101(.025)[.025]	.099(.025)[.025]	.796(.050)[.050]	.802(.045)[.045]
2SLS-2	.216(.037)[.040]	.099(.022)[.022]	.095(.021)[.022]	.793(.050)[.050]	.800(.043)[.043]
BC2SLS	.199(.039)[.039]	.100(.022)[.022]	.100(.022)[.022]	.796(.050)[.050]	.803(.043)[.044]
3SLS-1	.202(.039)[.039]	.101(.025)[.025]	.099(.025)[.025]	.797(.050)[.050]	.802(.045)[.045]
3SLS-2	.211(.038)[.040]	.099(.022)[.022]	.096(.022)[.022]	.795(.050)[.050]	.801(.043)[.043]
BC3SLS	.200(.039)[.039]	.100(.022)[.022]	.099(.022)[.022]	.797(.050)[.050]	.802(.044)[.044]
$\sigma_{12} = 0.5$					
2SLS-1	.202(.039)[.039]	.101(.025)[.025]	.099(.025)[.025]	.796(.050)[.050]	.802(.045)[.045]
2SLS-2	.235(.036)[.050]	.099(.021)[.021]	.088(.021)[.024]	.790(.049)[.050]	.794(.042)[.042]
BC2SLS	.196(.040)[.040]	.100(.022)[.022]	.100(.022)[.022]	.797(.050)[.050]	.804(.044)[.044]
3SLS-1	.203(.039)[.039]	.101(.025)[.025]	.098(.024)[.024]	.799(.047)[.047]	.800(.043)[.043]
3SLS-2	.212(.038)[.040]	.099(.021)[.021]	.095(.021)[.022]	.801(.046)[.046]	.803(.042)[.042]
BC3SLS	.201(.040)[.040]	.101(.022)[.022]	.099(.022)[.022]	.799(.047)[.047]	.801(.043)[.043]
$\sigma_{12} = 0.9$					
2SLS-1	.203(.039)[.039]	.101(.025)[.025]	.098(.024)[.025]	.796(.050)[.050]	.802(.045)[.045]
2SLS-2	.253(.034)[.063]	.100(.020)[.020]	.081(.020)[.027]	.787(.047)[.049]	.789(.041)[.042]
BC2SLS	.193(.042)[.042]	.101(.022)[.022]	.101(.022)[.022]	.797(.051)[.051]	.804(.044)[.044]
3SLS-1	.203(.039)[.039]	.102(.023)[.023]	.098(.024)[.024]	.802(.037)[.037]	.800(.037)[.037]
3SLS-2	.213(.041)[.043]	.098(.020)[.020]	.094(.021)[.022]	.807(.039)[.040]	.806(.039)[.039]
BC3SLS	.202(.042)[.042]	.102(.021)[.021]	.098(.022)[.022]	.801(.040)[.040]	.800(.040)[.040]

Mean(SD)[RMSE]

Table 3: 2SLS and 3SLS Estimation ($n_r = 30$, $\bar{r} = 30$)

	$\varphi_1 = 0.2$	$\lambda_{11} = 0.1$	$\lambda_{21} = 0.1$	$\beta_1 = 0.8$	$\gamma_1 = 0.8$
$\sigma_{12} = 0.1$					
2SLS-1	.200(.029)[.029]	.101(.016)[.016]	.099(.018)[.018]	.800(.034)[.034]	.799(.033)[.033]
2SLS-2	.207(.028)[.029]	.100(.015)[.015]	.097(.016)[.017]	.799(.034)[.034]	.798(.031)[.031]
BC2SLS	.199(.029)[.029]	.100(.015)[.015]	.100(.017)[.017]	.801(.034)[.034]	.800(.032)[.032]
3SLS-1	.200(.029)[.029]	.101(.016)[.016]	.099(.018)[.018]	.800(.034)[.034]	.799(.033)[.033]
3SLS-2	.205(.029)[.029]	.100(.015)[.015]	.098(.016)[.017]	.800(.034)[.034]	.799(.032)[.032]
BC3SLS	.199(.029)[.029]	.100(.015)[.015]	.100(.016)[.016]	.800(.034)[.034]	.800(.032)[.032]
$\sigma_{12} = 0.5$					
2SLS-1	.200(.029)[.029]	.101(.016)[.016]	.099(.018)[.018]	.800(.034)[.034]	.799(.033)[.033]
2SLS-2	.216(.028)[.032]	.100(.014)[.014]	.094(.016)[.017]	.798(.034)[.034]	.795(.031)[.031]
BC2SLS	.197(.029)[.029]	.100(.015)[.015]	.100(.017)[.017]	.801(.034)[.034]	.800(.032)[.032]
3SLS-1	.200(.029)[.029]	.100(.016)[.016]	.099(.018)[.018]	.800(.032)[.032]	.800(.031)[.031]
3SLS-2	.205(.028)[.029]	.099(.014)[.014]	.098(.016)[.016]	.801(.032)[.032]	.801(.030)[.030]
BC3SLS	.200(.029)[.029]	.100(.014)[.014]	.099(.016)[.016]	.800(.032)[.032]	.800(.030)[.030]
$\sigma_{12} = 0.9$					
2SLS-1	.201(.029)[.029]	.101(.016)[.016]	.099(.018)[.018]	.800(.034)[.034]	.799(.033)[.033]
2SLS-2	.225(.027)[.037]	.101(.014)[.014]	.090(.016)[.019]	.796(.033)[.033]	.793(.031)[.032]
BC2SLS	.196(.030)[.030]	.100(.015)[.015]	.101(.017)[.017]	.801(.034)[.034]	.801(.032)[.032]
3SLS-1	.201(.029)[.029]	.100(.015)[.015]	.099(.018)[.018]	.800(.027)[.027]	.800(.027)[.027]
3SLS-2	.205(.029)[.030]	.099(.014)[.014]	.098(.016)[.017]	.803(.027)[.027]	.803(.027)[.027]
BC3SLS	.200(.030)[.030]	.100(.014)[.014]	.099(.017)[.017]	.800(.028)[.028]	.800(.027)[.027]

Mean(SD)[RMSE]

Table 4: 2SLS and 3SLS Estimation ($n_r = 15$, $\bar{r} = 60$)

	$\varphi_1 = 0.2$	$\lambda_{11} = 0.1$	$\lambda_{21} = 0.1$	$\beta_1 = 0.8$	$\gamma_1 = 0.8$
$\sigma_{12} = 0.1$					
2SLS-1	.200(.029)[.029]	.098(.018)[.018]	.101(.020)[.020]	.798(.035)[.035]	.802(.032)[.032]
2SLS-2	.213(.028)[.031]	.096(.016)[.016]	.097(.017)[.018]	.795(.035)[.035]	.800(.031)[.031]
BC2SLS	.198(.030)[.030]	.099(.016)[.016]	.102(.018)[.018]	.799(.035)[.035]	.802(.031)[.031]
3SLS-1	.200(.029)[.029]	.098(.018)[.018]	.101(.020)[.020]	.798(.035)[.035]	.802(.032)[.032]
3SLS-2	.208(.030)[.031]	.096(.016)[.016]	.099(.018)[.018]	.797(.035)[.035]	.802(.031)[.031]
BC3SLS	.199(.030)[.030]	.099(.016)[.016]	.101(.018)[.018]	.798(.035)[.035]	.802(.031)[.031]
$\sigma_{12} = 0.5$					
2SLS-1	.200(.029)[.029]	.098(.018)[.018]	.101(.019)[.020]	.798(.035)[.035]	.802(.032)[.032]
2SLS-2	.231(.028)[.042]	.097(.016)[.016]	.091(.017)[.019]	.792(.034)[.035]	.795(.030)[.031]
BC2SLS	.196(.030)[.031]	.099(.016)[.016]	.102(.018)[.018]	.799(.035)[.035]	.802(.031)[.031]
3SLS-1	.200(.029)[.029]	.098(.018)[.018]	.101(.019)[.019]	.799(.033)[.033]	.802(.031)[.031]
3SLS-2	.209(.030)[.031]	.095(.016)[.016]	.098(.018)[.018]	.800(.033)[.033]	.805(.029)[.030]
BC3SLS	.199(.030)[.030]	.099(.016)[.016]	.101(.018)[.018]	.798(.033)[.033]	.801(.030)[.030]
$\sigma_{12} = 0.9$					
2SLS-1	.200(.029)[.029]	.098(.018)[.018]	.101(.019)[.019]	.798(.035)[.035]	.802(.032)[.032]
2SLS-2	.250(.027)[.056]	.097(.015)[.015]	.084(.017)[.023]	.789(.033)[.035]	.789(.029)[.031]
BC2SLS	.195(.031)[.032]	.099(.016)[.016]	.102(.018)[.019]	.799(.035)[.035]	.803(.031)[.032]
3SLS-1	.200(.029)[.029]	.098(.017)[.017]	.101(.019)[.019]	.800(.028)[.028]	.801(.027)[.027]
3SLS-2	.210(.031)[.032]	.095(.015)[.016]	.097(.018)[.018]	.805(.028)[.029]	.807(.027)[.028]
BC3SLS	.200(.031)[.031]	.099(.015)[.015]	.101(.018)[.018]	.799(.029)[.029]	.801(.027)[.027]

Mean(SD)[RMSE]

Table 5: 2SLS and 3SLS Estimation ($n_r = 10$, $\bar{r} = 30$)

	$\varphi_1 = 0.2$	$\lambda_{11} = 0.1$	$\lambda_{21} = 0.1$	$\beta_1 = 0.4$	$\gamma_1 = 0.4$
$\sigma_{12} = 0.1$					
2SLS-1	.210(.105)[.106]	.097(.072)[.072]	.100(.065)[.065]	.406(.067)[.067]	.405(.057)[.057]
2SLS-2	.272(.086)[.112]	.078(.045)[.050]	.092(.043)[.044]	.396(.063)[.063]	.402(.050)[.050]
BC2SLS	.194(.120)[.120]	.098(.050)[.050]	.104(.049)[.049]	.406(.065)[.066]	.405(.053)[.053]
3SLS-1	.210(.106)[.106]	.097(.072)[.072]	.100(.065)[.065]	.406(.067)[.067]	.405(.057)[.057]
3SLS-2	.253(.111)[.122]	.080(.046)[.050]	.096(.046)[.046]	.397(.064)[.064]	.403(.051)[.051]
BC3SLS	.199(.112)[.112]	.098(.050)[.050]	.103(.048)[.048]	.406(.065)[.066]	.404(.053)[.053]
$\sigma_{12} = 0.5$					
2SLS-1	.213(.104)[.105]	.098(.071)[.071]	.098(.065)[.065]	.406(.066)[.067]	.404(.056)[.056]
2SLS-2	.360(.078)[.178]	.078(.042)[.047]	.069(.041)[.051]	.389(.058)[.059]	.390(.046)[.047]
BC2SLS	.178(.126)[.128]	.099(.051)[.051]	.108(.051)[.052]	.407(.066)[.067]	.407(.054)[.054]
3SLS-1	.213(.105)[.106]	.097(.069)[.069]	.097(.064)[.065]	.407(.062)[.062]	.406(.054)[.054]
3SLS-2	.258(.104)[.119]	.080(.045)[.049]	.091(.045)[.046]	.406(.059)[.059]	.410(.047)[.048]
BC3SLS	.204(.111)[.111]	.098(.050)[.050]	.102(.049)[.049]	.406(.061)[.062]	.403(.051)[.051]
$\sigma_{12} = 0.9$					
2SLS-1	.216(.103)[.104]	.099(.070)[.070]	.096(.065)[.065]	.406(.066)[.066]	.404(.055)[.056]
2SLS-2	.441(.065)[.250]	.081(.035)[.040]	.046(.036)[.065]	.384(.047)[.050]	.378(.039)[.045]
BC2SLS	.167(.145)[.149]	.100(.055)[.055]	.111(.056)[.057]	.409(.068)[.069]	.409(.056)[.056]
3SLS-1	.215(.104)[.105]	.095(.063)[.063]	.096(.064)[.064]	.409(.049)[.050]	.409(.046)[.047]
3SLS-2	.264(.100)[.119]	.081(.044)[.048]	.086(.045)[.047]	.421(.051)[.055]	.422(.047)[.052]
BC3SLS	.214(.111)[.112]	.098(.050)[.050]	.102(.051)[.051]	.409(.052)[.053]	.408(.048)[.049]

Mean(SD)[RMSE]

Table 6: 2SLS and 3SLS Estimation ($n_r = 15$, $\bar{r} = 30$)

	$\varphi_1 = 0.2$	$\lambda_{11} = 0.1$	$\lambda_{21} = 0.1$	$\beta_1 = 0.4$	$\gamma_1 = 0.4$
$\sigma_{12} = 0.1$					
2SLS-1	.204(.078)[.078]	.100(.052)[.052]	.098(.050)[.050]	.396(.051)[.051]	.403(.045)[.046]
2SLS-2	.250(.067)[.083]	.095(.034)[.034]	.087(.034)[.036]	.392(.049)[.050]	.398(.040)[.040]
BC2SLS	.194(.084)[.084]	.100(.035)[.035]	.101(.037)[.037]	.397(.050)[.050]	.404(.042)[.042]
3SLS-1	.204(.078)[.078]	.100(.052)[.052]	.098(.050)[.050]	.397(.050)[.051]	.403(.045)[.046]
3SLS-2	.236(.078)[.086]	.096(.034)[.034]	.090(.035)[.037]	.394(.049)[.049]	.400(.041)[.041]
BC3SLS	.198(.081)[.081]	.101(.035)[.035]	.100(.036)[.036]	.398(.050)[.050]	.403(.042)[.042]
$\sigma_{12} = 0.5$					
2SLS-1	.206(.078)[.078]	.100(.051)[.051]	.097(.049)[.050]	.396(.051)[.051]	.403(.045)[.045]
2SLS-2	.311(.061)[.126]	.095(.032)[.032]	.070(.032)[.044]	.388(.046)[.048]	.390(.038)[.039]
BC2SLS	.184(.088)[.090]	.101(.036)[.036]	.103(.038)[.038]	.397(.051)[.051]	.405(.043)[.043]
3SLS-1	.206(.078)[.078]	.102(.050)[.050]	.096(.049)[.049]	.400(.047)[.047]	.402(.043)[.043]
3SLS-2	.239(.077)[.086]	.095(.034)[.034]	.088(.034)[.036]	.403(.045)[.045]	.405(.040)[.040]
BC3SLS	.201(.083)[.083]	.102(.035)[.035]	.098(.036)[.037]	.399(.047)[.047]	.401(.042)[.042]
$\sigma_{12} = 0.9$					
2SLS-1	.207(.078)[.079]	.101(.051)[.051]	.096(.049)[.050]	.396(.051)[.051]	.403(.045)[.045]
2SLS-2	.368(.052)[.176]	.095(.028)[.029]	.053(.029)[.055]	.384(.041)[.044]	.382(.034)[.039]
BC2SLS	.173(.096)[.100]	.103(.037)[.037]	.104(.039)[.039]	.398(.052)[.052]	.406(.044)[.044]
3SLS-1	.208(.078)[.078]	.102(.047)[.047]	.094(.048)[.048]	.404(.037)[.038]	.402(.037)[.037]
3SLS-2	.241(.080)[.090]	.094(.033)[.033]	.087(.034)[.036]	.413(.038)[.040]	.412(.038)[.039]
BC3SLS	.204(.084)[.084]	.102(.034)[.035]	.096(.036)[.037]	.403(.039)[.040]	.402(.039)[.039]

Mean(SD)[RMSE]

Table 7: 2SLS and 3SLS Estimation ($n_r = 30$, $\bar{r} = 30$)

	$\varphi_1 = 0.2$	$\lambda_{11} = 0.1$	$\lambda_{21} = 0.1$	$\beta_1 = 0.4$	$\gamma_1 = 0.4$
$\sigma_{12} = 0.1$					
2SLS-1	.200(.058)[.058]	.101(.033)[.033]	.099(.037)[.037]	.400(.034)[.034]	.400(.033)[.033]
2SLS-2	.224(.052)[.057]	.099(.024)[.024]	.092(.026)[.028]	.398(.034)[.034]	.397(.029)[.030]
BC2SLS	.195(.058)[.058]	.100(.024)[.024]	.101(.028)[.028]	.401(.034)[.034]	.401(.030)[.030]
3SLS-1	.200(.058)[.058]	.101(.033)[.033]	.099(.037)[.037]	.400(.034)[.034]	.400(.033)[.033]
3SLS-2	.216(.056)[.059]	.099(.024)[.024]	.095(.027)[.028]	.399(.034)[.034]	.399(.030)[.030]
BC3SLS	.197(.057)[.057]	.100(.024)[.024]	.100(.028)[.028]	.401(.034)[.034]	.400(.030)[.030]
$\sigma_{12} = 0.5$					
2SLS-1	.201(.058)[.058]	.101(.033)[.033]	.098(.037)[.037]	.400(.034)[.034]	.400(.033)[.033]
2SLS-2	.256(.050)[.074]	.099(.023)[.023]	.083(.026)[.031]	.396(.033)[.033]	.393(.029)[.029]
BC2SLS	.189(.060)[.061]	.101(.024)[.024]	.102(.029)[.029]	.401(.034)[.034]	.401(.030)[.030]
3SLS-1	.201(.058)[.058]	.100(.032)[.032]	.098(.036)[.036]	.400(.032)[.032]	.400(.031)[.031]
3SLS-2	.217(.056)[.058]	.098(.023)[.023]	.094(.027)[.028]	.402(.031)[.031]	.402(.028)[.028]
BC3SLS	.198(.058)[.058]	.100(.024)[.024]	.099(.028)[.028]	.400(.032)[.032]	.400(.029)[.029]
$\sigma_{12} = 0.9$					
2SLS-1	.202(.058)[.058]	.101(.033)[.033]	.098(.037)[.037]	.400(.034)[.034]	.399(.033)[.033]
2SLS-2	.286(.046)[.098]	.099(.022)[.022]	.073(.024)[.036]	.393(.031)[.032]	.389(.027)[.029]
BC2SLS	.183(.063)[.065]	.101(.025)[.025]	.104(.029)[.030]	.402(.035)[.035]	.402(.031)[.031]
3SLS-1	.202(.057)[.057]	.100(.030)[.030]	.098(.036)[.036]	.401(.027)[.027]	.401(.027)[.027]
3SLS-2	.218(.058)[.060]	.096(.023)[.023]	.095(.027)[.028]	.406(.027)[.028]	.406(.026)[.027]
BC3SLS	.199(.060)[.060]	.100(.023)[.023]	.099(.029)[.029]	.401(.028)[.028]	.401(.027)[.027]

Mean(SD)[RMSE]

Table 8: 2SLS and 3SLS Estimation ($n_r = 15$, $\bar{r} = 60$)

	$\varphi_1 = 0.2$	$\lambda_{11} = 0.1$	$\lambda_{21} = 0.1$	$\beta_1 = 0.4$	$\gamma_1 = 0.4$
$\sigma_{12} = 0.1$					
2SLS-1	.199(.059)[.059]	.096(.036)[.036]	.103(.039)[.039]	.398(.035)[.035]	.402(.032)[.032]
2SLS-2	.244(.050)[.067]	.091(.025)[.026]	.092(.027)[.028]	.394(.035)[.035]	.398(.028)[.028]
BC2SLS	.193(.062)[.062]	.099(.025)[.025]	.103(.029)[.030]	.399(.035)[.035]	.402(.029)[.030]
3SLS-1	.199(.059)[.059]	.096(.036)[.036]	.103(.039)[.039]	.398(.035)[.035]	.402(.032)[.032]
3SLS-2	.228(.060)[.066]	.091(.025)[.026]	.096(.029)[.029]	.396(.035)[.035]	.401(.029)[.029]
BC3SLS	.195(.060)[.060]	.099(.025)[.025]	.102(.029)[.029]	.399(.035)[.035]	.401(.029)[.029]
$\sigma_{12} = 0.5$					
2SLS-1	.200(.059)[.059]	.096(.036)[.036]	.103(.039)[.039]	.398(.035)[.035]	.402(.032)[.032]
2SLS-2	.303(.047)[.114]	.090(.024)[.025]	.075(.026)[.036]	.389(.032)[.034]	.390(.027)[.029]
BC2SLS	.187(.065)[.066]	.099(.026)[.026]	.104(.031)[.031]	.399(.035)[.035]	.402(.030)[.030]
3SLS-1	.200(.059)[.059]	.096(.035)[.036]	.103(.039)[.039]	.399(.033)[.033]	.402(.031)[.031]
3SLS-2	.230(.060)[.067]	.090(.025)[.027]	.094(.029)[.029]	.401(.033)[.033]	.406(.027)[.028]
BC3SLS	.197(.063)[.063]	.099(.026)[.026]	.102(.030)[.030]	.398(.034)[.034]	.401(.029)[.029]
$\sigma_{12} = 0.9$					
2SLS-1	.200(.059)[.059]	.096(.036)[.036]	.102(.039)[.039]	.398(.035)[.035]	.402(.032)[.033]
2SLS-2	.360(.041)[.165]	.092(.021)[.023]	.058(.024)[.048]	.385(.029)[.032]	.382(.024)[.030]
BC2SLS	.181(.069)[.072]	.099(.026)[.026]	.106(.032)[.033]	.400(.036)[.036]	.403(.031)[.031]
3SLS-1	.201(.058)[.058]	.096(.033)[.034]	.102(.038)[.038]	.401(.028)[.028]	.402(.027)[.027]
3SLS-2	.232(.062)[.069]	.090(.025)[.027]	.092(.029)[.030]	.409(.028)[.030]	.411(.026)[.028]
BC3SLS	.198(.064)[.064]	.099(.026)[.026]	.101(.031)[.031]	.400(.029)[.029]	.401(.027)[.027]

Mean(SD)[RMSE]

Table 9: Data Summary

Variable	Definition	Mean	SD
Delinquency	In the text	1.25	0.70
Sports	In the text	1.16	1.04
Age	Age	15.47	1.68
Female	1 if the respondent is female	0.54	0.50
(White)	1 if the respondent is white	0.62	0.49
African American	1 if the respondent is African American	0.18	0.38
Asian American	1 if the respondent is Asian American	0.05	0.22
Other race	1 if the respondent is not white or African American	0.15	0.36
Grade	Grade, coded as 1 if the respondent is in grade 6, 2 if the respondent is in grade 7, etc.	3.55	1.62
Both parents	1 if the respondent lives with both parents	0.75	0.43
Parental care	1 if the respondent thinks parents care about him/her very much	0.85	0.36
(Less than HS)	1 if the parent's education is less than high school (HS)	0.12	0.33
HS graduate	1 if the parent's education is HS or higher but no college degree	0.54	0.50
College graduate	1 if the parent's education is college or higher	0.29	0.45
Education missing	1 if the parent's education information is missing	0.04	0.20
Professional	1 if the parent's job is a doctor, lawyer, scientist, teacher, librarian, nurse, manager, executive, director, computer specialist, and radiologist	0.29	0.45
Office worker	1 if the parent's job is office worker, bookkeeper, clerk, secretary, sales worker, insurance agent and store clerk	0.23	0.42
Other job	1 if the parent's job is not listed above	0.30	0.46
(Unemployed)	1 if the parent is unemployed	0.12	0.33
Job missing	1 if the parent's job information is missing	0.05	0.21
Living condition	1 if the building in which the respondent lives is very well kept	0.61	0.49
Parent smoking	1 if the parent of the respondent smokes cigarettes	0.64	0.48
Parent religious	1 if the parent went to a religious service with the respondent in the past 4 weeks	0.43	0.49
Parent sports	1 if the parent played a sport with the respondent in the past 4 weeks	0.27	0.45
Coordination	1 if the respondent is well coordinated	0.35	0.48

The variable in the parentheses is the reference category.

If both parents are in the household, the education and job of the father is considered.

Table 10: Estimation of Peer Effects in Juvenile Delinquency

	2SLS	BC2SLS	3SLS	BC3SLS
Simultaneity effect φ_1	-0.2916***	-0.3746***	-0.3666***	-0.2738***
	(0.0878)	(0.0562)	(0.0865)	(0.0488)
Endogenous peer effect λ_{11}	0.0590***	0.0376***	0.0573***	0.0355***
	(0.0152)	(0.0126)	(0.0151)	(0.0122)
Cross-activity peer effect λ_{12}	-0.0448***	-0.0272**	-0.0424***	-0.0264**
	(0.0141)	(0.0118)	(0.0140)	(0.0115)
Age	-0.0024	-0.0026	-0.0039	0.0015
	(0.0217)	(0.0219)	(0.0220)	(0.0215)
Female	-0.4333***	-0.4567***	-0.4533***	-0.4289***
	(0.0360)	(0.0313)	(0.0360)	(0.0299)
African American	-0.3201***	-0.3294***	-0.3306***	-0.3283***
	(0.0603)	(0.0610)	(0.0610)	(0.0599)
Asian American	-0.4782***	-0.4907***	-0.4864***	-0.4835***
	(0.0850)	(0.0860)	(0.0861)	(0.0846)
Other race	-0.1048**	-0.1009**	-0.1057**	-0.1061**
	(0.0463)	(0.0469)	(0.0469)	(0.0461)
Grade	-0.0100	-0.0144	-0.0152	-0.0099
	(0.0266)	(0.0267)	(0.0269)	(0.0262)
Both parents	-0.0097	-0.0074	-0.0054	-0.0092
	(0.0336)	(0.0340)	(0.0339)	(0.0333)
Parental care	-0.4632***	-0.4639***	-0.4644***	-0.4668***
	(0.0386)	(0.0391)	(0.0391)	(0.0385)
HS graduate	-0.0145	-0.0144	-0.0152	-0.0228
	(0.0441)	(0.0446)	(0.0447)	(0.0438)
College graduate	0.0263	0.0262	0.0225	0.0116
	(0.0517)	(0.0521)	(0.0523)	(0.0510)
Education missing	-0.0726	-0.0817	-0.0880	-0.0838
	(0.0788)	(0.0798)	(0.0796)	(0.0783)
Professional	0.0104	0.0124	0.0112	0.0084
	(0.0475)	(0.0482)	(0.0481)	(0.0474)
Office worker	0.0355	0.0350	0.0336	0.0360
	(0.0480)	(0.0486)	(0.0486)	(0.0478)
Other job	-0.0321	-0.0348	-0.0338	-0.0311
	(0.0459)	(0.0464)	(0.0464)	(0.0457)
Job missing	0.0894	0.0922	0.0844	0.0879
	(0.0758)	(0.0769)	(0.0767)	(0.0755)
Living condition	-0.0492*	-0.0492*	-0.0516*	-0.0505*
	(0.0289)	(0.0293)	(0.0292)	(0.0288)
Parent smoking	0.1737***	0.1756***	0.1330***	0.1516***
	(0.0289)	(0.0292)	(0.0237)	(0.0240)
Parent religious	-0.1382***	-0.1370***	-0.1286***	-0.1494***
	(0.0297)	(0.0298)	(0.0236)	(0.0244)
First Stage F test statistic	11.590			
OIR test p-value	0.470			
J test p-value	0.823			

Standard errors in parentheses. Statistical significance: ***p<0.01; **p<0.05; *p<0.1.
Contextual effects and network fixed effects are considered for all estimators.