# Endogenous Peer Effects: Local Aggregate or Local Average?*

Xiaodong Liu[†] Eleonora Patacchini[‡] Yves Zenou[§]

October 8, 2013 (Revised: March 21, 2014)

**Abstract**

We develop a unified model embedding different behavioral mechanisms of social interactions and design a statistical model selection test to differentiate between them in empirical applications. This framework is applied to study peer effects in education (effort in studying) and sport activities for adolescents in the United States. We find that, for education, students tend to *conform* to the social norm of their friends while, for sport activities, both the social multiplier and the social norm effect matter.

*Key words:* Social networks, identification of peer effects, J test, individual-based policy, group-based policy.

*JEL classification:* A14, D85, Z13

*We thank the editor, William Neilson, as well as three anonymous referees for very helpful comments. We also that Maria Marta Ferreyra, Benjamin Golub, Lung-fei Lee and Marc Sommer for comments and discussions. We are also grateful to Edoardo Rainone for excellent research assistance. This research uses data from Add Health, a program project directed by Kathleen Mullan Harris and designed by J. Richard Udry, Peter S. Bearman and Kathleen Mullan Harris at the University of North Carolina at Chapel Hill, and funded by grant P01-HD31921 from the Eunice Kennedy Shriver National Institute of Child Health and Human Development, with cooperative funding from 23 other federal agencies and foundations. Special acknowledgment is due to Ronald R. Rindfuss and Barbara Entwisle for assistance in the original design. Information on how to obtain the Add Health data files is available on the Add Health website (http://www.cpc.unc.edu/addhealth). No direct support was received from grant P01-HD31921 for this analysis.

[†]University of Colorado Boulder. E-mail: xiaodong.liu@colorado.edu.

[‡]Syracuse University, EIEF and CEPR. E-mail: epatacch@maxwell.syr.edu

[§]Corresponding author. Stockholm University, Research Institute of Industrial Economics (IFN) and GAINS. E-mail: yves.zenou@ne.su.se.

## 1 Introduction

In many circumstances, the decision of agents to exert effort in education, or some other activity, cannot adequately be explained by their characteristics and by the intrinsic utility derived from it. Rather, its rationale may be found in how peers and others value this activity. There is indeed strong evidence that the behavior of individual agents is affected by that of their peers. This is particularly true in education, crime, labor markets, fertility, participation in welfare programs, etc. (for surveys, see, Glaeser and Scheinkman, 2001; Moffitt, 2001; Durlauf, 2004; Ioannides and Loury, 2004; Ioannides, 2012). The way peer effects operate is, however, unclear. Are students working hard at school because some of their friends work hard or because they do not want to be different from the majority of their peers who work hard?

The aim of this paper is to help our understanding of social interaction mechanisms of peer effects. For that, we begin by developing a *social network model* aiming at capturing how peer effects operate through social networks.[1] We characterize the Nash equilibrium and show under which condition an interior Nash equilibrium exists and is unique. Such a model encompasses the most popular peer effects models on networks: the *local-aggregate* and the *local-average* models. In the *local-aggregate model* (see, in particular, Ballester et al., 2006, 2010; Bramoullé and Kranton, 2007; Galeotti et al., 2009; Calvó-Armengol et al., 2009), endogenous peer effects are captured by the *sum of friends' efforts* in some activity so that the more active friends an individual has, the higher is her marginal utility of exerting effort. In the *local-average model* (e.g. Glaeser and Scheinkman, 2003; Patacchini and Zenou, 2012; Boucher et al., 2014), peers' choices are viewed as a *social norm* and individuals pay a cost for deviating from this norm. In this model, each individual wants to conform as much as possible to the social norm of her reference group, which is defined as the average effort of her friends.[2] Ghiglino and Goyal (2010) develop a theoretical model where they compare the local aggregate and local average models in the context of a pure exchange economy where individuals trade in markets and are influenced by their neighbors. They found that with *aggregate* comparisons, networks matter even if all people are equally wealthy. With *average* comparisons,

[1] There is a growing literature on networks in economics. See the recent literature surveys by Goyal (2007), Jackson (2008) and Jackson and Zenou (2013).

[2] In economics, different aspects of conformism and social norms have been explored from a theoretical point of view. To name a few, (i) peer pressures and partnerships (Kandel and Lazear, 1992) where peer pressure arises when individuals deviate from a well-established group norm, e.g. individuals are penalized for working less than the group norm, (ii) religion (Iannaccone 1992, Berman 2000), since praying is much more satisfying the more average participants there are, (iii) social status and social distance (Akerlof 1980, 1997; Bernheim 1994; Battu et al., 2007, among others) where deviations from the social norm (average action) imply a loss of reputation and status.

networks are irrelevant when individuals are equally wealthy. The two models are, however, similar if there is heterogeneity in wealth.[3] We are not aware of a paper where both local-aggregate and local-average effects are incorporated in a unified network model.

Next, we study the econometric counterpart of the theoretical model. In the spatial econometric literature, the local-average and the local-aggregate model are well-known and their main difference (from an econometric viewpoint) is due to the fact that the adjacency matrix is row-normalized in the former but not in the latter. Our theoretical analysis provides a microfoundation for these two models. For the *local-average* model, Bramoullé et al. (2009) show that intransitivity in network connections can be used as an exclusion restriction to identify the endogenous peer effect from contextual and correlated effects. In this paper, we show that, for the *local-aggregate* model, different positions of the agents in a network captured by the Bonacich (1987) centrality can be used as additional instruments to improve identification and estimation efficiency. We also give identification conditions for a general econometric network model that incorporates both local-aggregate and local-average endogenous peer effects.

Finally, we extend Kelejian's (2008) J test for spatial econometric models to differentiate between the local-aggregate and the local-average endogenous peer effects in an econometric network model with network fixed-effects. We illustrate our methodology using data from the U.S. National Longitudinal Survey of Adolescent Health (AddHealth), which contains unique detailed information on friendship relationships among teenagers. In line with a number of recent studies based on the AddHealth data (e.g. Calvó-Armengol et al., 2009, Lin 2010; Patacchini and Zenou, 2012; Liu et al. 2012), we exploit the structure of the network as well as network fixed effects to identify peer effects from contextual and correlated effects.[4] We find that, for study effort, students tend to *conform* to the social norm of their friends while, for sport activities, both the social multiplier and the social norm effect matter. Our results also show that the local-average peer effect is overstated if the local-aggregate effect is ignored and vice versa. In this respect, our analysis reveals that caution is warranted in the assessment of peer effects when social interactions can take different forms.

We believe that it is important to be able to disentangle empirically different behavioral mechanisms of endogenous peer effects because they imply different policy implications. In the *local-average*

[3] Another interesting paper is that of Clark and Oswald (1998) who propose a choice-theoretical justification for the local-average (i.e. conformist) model.

[4] The underlying assumption is that any troubling source of heterogeneity, which is left unexplained by the set of observed (individual and peers) characteristics can be captured at the network level, and thus taken into account by the inclusion of network fixed effects.

model, the only way to affect individuals' behavior and thus their outcomes is to change the *social norm* of the group. In other words, one needs to affect most people in the group for the policy to be effective. As a result, *group-based policies* should be implemented in the context of this model. On the other hand, for the *local-aggregate* model, one can target only one individual and still effectively influence the whole network. In an other words, in the local-aggregate model there is a more salient social multiplier effect than in the local-average model, and hence, individual-based policies could be implemented.[5]

The rest of paper is organized as follows. Section 2 introduces the theoretical framework for the network models. Section 3 discusses the identification conditions of the corresponding econometric models. We extend the J test of Kelejian and Piras (2011) to network models with network fixed effects in Section 4 and empirically test the network models using the AddHealth data in Section 5. Section 5.4 discusses the policy implications of our results. Finally, Section 6 concludes. All proofs of propositions can be found in Appendix A.

# 2 Theoretical Framework

## 2.1 The network

Suppose that a finite set of agents $N = \{1, \ldots, n\}$ is partitioned into $\bar{r}$ networks, where $N_r = \{1, \ldots, n_r\}$ denotes the set of agents in the $r$-th network ($r = 1, ..., \bar{r}$).[6] We keep track of social connections in network $r$ by its *adjacency matrix* $G_r = [g_{ij,r}]$, where $g_{ij,r} = 1$ if $i$ and $j$ are friends, and $g_{ij,r} = 0$, otherwise.[7] We also set $g_{ii,r} = 0$.

The *reference group* of individual $i$ in network $r$ is the set of $i$'s friends given by $N_{i,r} = \{j \neq i \mid g_{ij,r} = 1\}$. The size of $N_{i,r}$ is $g_{i,r} = \sum_{j=1}^{n_r} g_{ij,r}$, which is known as the *degree* of $i$ in graph theory.[8]

Let $G_r^* = [g_{ij,r}^*]$, where $g_{ij,r}^* = g_{ij,r}/g_{i,r}$, denote the *row-normalized* adjacency matrix of network $r$. By construction, we have $0 \leq g_{ij,r}^* \leq 1$ and $\sum_{j=1}^{n_r} g_{ij,r}^* = 1$. Figure 1 gives an example of a star-

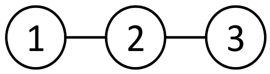
[5] See our discussion in Sections 2.5 and 6.

[6] In the theoretical analysis, we will only consider *one* connected network (i.e. there always exists a path between two individuals in this network). We keep, however, the notation of $r$ (multiple) networks (each being connected) to ease the transition between the theoretical and the econometric analysis.

[7] We assume friendships are reciprocal so that $g_{ij,r} = g_{ji,r}$. All our results hold for asymmetric (*directed*) and *weighted* networks but, for the ease of the presentation, we focus on symmetric (*undirected*) and *unweighted* networks in this paper.

[8] For simplicity, we assume that no one is isolated so that $g_{i,r} > 0$.

shaped network and the corresponding adjacency matrices with and without row-normalization.




$$G_r = \begin{bmatrix} 0 & 1 & 0 \\ 1 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix} \quad \text{and} \quad G_r^* = \begin{bmatrix} 0 & 1 & 0 \\ 1/2 & 0 & 1/2 \\ 0 & 1 & 0 \end{bmatrix}.$$

Figure 1: an example network with corresponding adjacency matrices

### 2.2 The general network model

We denote by $y_{i,r}$ the effort level of individual $i$ and by $Y_r = (y_{1,r}, ..., y_{n_r,r})'$ the population effort profile of network $r$. Given the underlying network topology represented by the adjacency matrix $G_r$, individuals in network $r$ simultaneously decide how much effort to exert to maximize the following utility function

$$u_{i,r}(y_{i,r}) \equiv u_{i,r}(y_{i,r}; Y_r, G_r) = \underbrace{(\pi^*_{i,r} + \lambda_1 \textstyle\sum_{j=1}^{n_r} g_{ij,r} y_{j,r}) y_{i,r}}_{\text{benefit}} - \underbrace{\frac{1}{2}[y^2_{i,r} + \lambda_2 (y_{i,r} - \textstyle\sum_{j=1}^{n_r} g^*_{ij,r} y_{j,r})^2]}_{\text{cost}}, \tag{1}$$

where $\lambda_1 \geq 0$ and $\lambda_2 \geq 0$. The utility function (1) has two components, benefit and cost. The benefit component is modeled as a linear function of own effort level $y_{i,r}$, with the term $(\pi^*_{i,r} + \lambda_1 \sum_{j=1}^{n_r} g_{ij,r} y_{j,r})$ representing the return to effort. The term $\pi^*_{i,r}$ represents *ex ante* individual heterogeneity in the return to effort. The term $\sum_{j=1}^{n_r} g_{ij,r} y_{j,r}$ represents the *aggregate* effort of $i$'s friends with the *social-multiplier* coefficient $\lambda_1$. As individuals may have different locations in the network, $\sum_{j=1}^{n_r} g_{ij,r} y_{j,r}$ is heterogeneous in $i$ even if every individual in the network chooses the same effort level.

The cost component of the utility function has two terms. The first term $y^2_{i,r}$ represents the cost of own effort. The second term $(y_{i,r} - \sum_{j=1}^{n_r} g^*_{ij,r} y_{j,r})^2$ represents the moral cost due to deviation from the *social norm* of the reference group (i.e., the average effort of the peers) with the *social-conformity* coefficient $\lambda_2$.[9]

Thus, an individual's utility is positively affected by the total effort of her friends *and* negatively affected by the distance from the average effort of her friends. From the first-order condition of the

[9] This is the standard way economists have been modeling conformity (see, among others, Akerlof, 1980, Bernheim, 1994, Kandel and Lazear, 1992, Akerlof, 1997, Fershtman and Weiss, 1998; Patacchini and Zenou, 2012).

utility maximization, the best-reply function of individual $i$ is then given by:

$$y_{i,r} = \phi_1 \sum_{j=1}^{n_r} g_{ij,r} y_{j,r} + \phi_2 \sum_{j=1}^{n_r} g^*_{ij,r} y_{j,r} + \pi_{i,r}, \tag{2}$$

where $\phi_1 = \lambda_1/(1+\lambda_2)$, $\phi_2 = \lambda_2/(1+\lambda_2)$, and $\pi_{i,r} = \pi^*_{i,r}/(1+\lambda_2)$. As $\lambda_1 \geq 0$ and $\lambda_2 \geq 0$, we have $\phi_1 \geq 0$ and $0 \leq \phi_2 < 1$. The coefficient $\phi_1$ is called the *local-aggregate* endogenous peer effect. As $\phi_1 \geq 0$, this coefficient reflects *strategic complementarity* in efforts. The coefficient $\phi_2$ is called the *local-average* endogenous peer effect, which captures the *taste for conformity.* Note that, $\phi_1/\phi_2 = \lambda_1/\lambda_2$. That is, the relative magnitude of $\phi_1$ and $\phi_2$ is the same as that of the social-multiplier coefficient $\lambda_1$ and the social-conformity coefficient $\lambda_2$.

We denote by $g_r^{\max}$ the highest degree in network $r$, i.e. $g_r^{\max} = \max_i g_{i,r}$. Let $\Pi_r = (\pi_{1,r}, \cdots, \pi_{n_r,r})'$. The Nash equilibrium of the general network model is characterized by the following proposition.

**Proposition 1** *If $\phi_1 \geq 0$, $\phi_2 \geq 0$ and $g_r^{\max}\phi_1 + \phi_2 < 1$, then the network game with payoffs (1) has a unique interior Nash equilibrium in pure strategies given by*

$$Y_r = (I_{n_r} - \phi_1 G_r - \phi_2 G_r^*)^{-1} \Pi_r. \tag{3}$$

We have a game with strategic complementarities (also called supermodular game) so that the higher is the effort of my neighbors, the higher is my marginal utility of exerting my own effort. As a result, there is a problem of existence of equilibrium because there is no bound on the effort of each agent. Since $\phi_1$ and $\phi_2$ express the intensity of these complementarities, then condition $g_r^{\max}\phi_1 + \phi_2 < 1$ guarantees the existence of equilibrium by limiting the degree of strategic complementarities of efforts. The fact that we do not have multiple equilibria is due to the fact that we have both linear best-reply functions and strategic complementarities. Indeed, introducing enough non-linearity in the best-replies, or enough substitutabilities in the interactions, would lead to multiple equilibria. In a way, linearity and strategic complementarities discipline each other.

### 2.3 Two special cases: local-average and local-aggregate network models

There are two special cases of the general network model that are of special interest. When $\lambda_1 = 0$, i.e., when there is no strategic complementarity effect, the utility function (1) reduces to

$$u_{i,r}(y_{i,r}) \equiv u_{i,r}(y_{i,r}; Y_r, G_r) = \underbrace{\pi^*_{i,r} y_{i,r}}_{\text{benefit}} - \underbrace{\frac{1}{2}[y^2_{i,r} + \lambda_2 (y_{i,r} - \textstyle\sum_{j=1}^{n_r} g^*_{ij,r} y_{j,r})^2]}_{\text{cost}}.$$

In equilibrium, agents choose their effort level $y_{i,r}$ simultaneously to maximize their utility. It follows from the first-order condition that the best-reply function of individual $i$ is given by

$$y_{i,r} = \phi_2 \textstyle\sum_{j=1}^{n_r} g^*_{ij,r} y_{j,r} + \pi_{i,r}, \tag{4}$$

where $\phi_2 = \lambda_2/(1 + \lambda_2)$ and $\pi_{i,r} = \pi^*_{i,r}/(1 + \lambda_2)$. As the equilibrium effort level of individual $i$ only depends on the *average* effort of her friends, this model is referred to as the *local-average* network model. If $0 \leq \phi_2 < 1$, then the local-average network game has a unique interior Nash equilibrium in pure strategies given by

$$Y_r = (I_{n_r} - \phi_2 G^*_r)^{-1} \Pi_r. \tag{5}$$

On the other hand, when $\lambda_2 = 0$, i.e., when there is no social-conformity effect, the utility function (1) reduces to

$$u_{i,r}(y_{i,r}) \equiv u_{i,r}(y_{i,r}; Y_r, G_r) = \underbrace{(\pi^*_{i,r} + \lambda_1 \textstyle\sum_{j=1}^{n_r} g_{ij,r} y_{j,r}) y_{i,r}}_{\text{benefit}} - \underbrace{\frac{1}{2} y^2_{i,r}}_{\text{cost}}.$$

In this case, it follows from the first-order condition that the best-reply function of individual $i$ is given by

$$y_{i,r} = \phi_1 \textstyle\sum_{j=1}^{n_r} g_{ij,r} y_{j,r} + \pi_{i,r}, \tag{6}$$

where $\phi_1 = \lambda_1/(1 + \lambda_2) = \lambda_1$ and $\pi_{i,r} = \pi^*_{i,r}/(1 + \lambda_2) = \pi^*_{i,r}$ (as $\lambda_2 = 0$). As the equilibrium effort level of individual $i$ depends on the *aggregate* effort of her friends. Hence, we call this model the *local-aggregate* network model. If $0 \leq g^{\max}_r \phi_1 < 1$, then the local-aggregate network game has a unique interior Nash equilibrium in pure strategies given by

$$Y_r = (I_{n_r} - \phi_1 G_r)^{-1} \Pi_r. \tag{7}$$

### 2.4 Equilibrium comparison: local average versus local aggregate

In the *local-average* model, it is the deviation from the *average effort* of the reference group that affects an individual's utility. So the closer an individual's effort is from the average effort of her friends, the higher is her equilibrium utility. On the contrary, in the local-aggregate model, it is the *aggregate effort* of the reference group that affects an individual's utility. So the more active friends an individual has, the higher is her equilibrium utility. In the *local-average* model, positions in the network would not matter and equilibrium effort would be the same if all individuals are *ex ante* identical. However, in the *local-aggregate* model, even if individuals are *ex ante* identical, different positions in the network would imply different equilibrium effort levels.

To illustrate this point, consider the case where all individuals are *ex ante* identical apart from their positions in the network such that $\pi_{i,r} = \pi_r$ for $i = 1, \cdots, n_r$. For the local-average model, if $0 \leq \phi_2 < 1$, the unique interior Nash equilibrium given by (5) now becomes

$$Y_r = \pi_r(1 - \phi_2)^{-1} l_{n_r},$$

where $l_{n_r}$ is an $n_r$-dimensional vector of ones. As a result, in the local-average model, the position in the network plays no role and all individuals provide the same equilibrium effort level $\pi_r/(1-\phi_2)$ in network $r$. On the other hand, for the local-aggregate model, if $0 \leq \phi_1 g_r^{\max} < 1$, the unique Nash equilibrium given by (7) now becomes

$$Y_r = \pi_r \left(I_{n_r} - \phi_1 G_r\right)^{-1} l_{n_r}.$$

Note that $\left(I_{n_r} - \phi_1 G_r\right)^{-1} l_{n_r}$ represents the Bonacich centrality (Bonacich, 1987) of a network. Therefore, the equilibrium effort $y_{i,r}$ of individual $i$ is proportional to her centrality in the network. The more central an individual's position is, the higher is her equilibrium effort and equilibrium utility.[10] This is one of the fundamental differences with the local-aggregate model where, even if agents are ex ante identical, because of social multiplier effects, the position in the network determines their effort activity so that more central persons exert more effort than less central individuals.

Thus, these two models have fundamentally different equilibrium implications as illustrated in

[10] For the local-aggregate model, the equilibrium utility of an individual is $u_{i,r}(y^*_{i,r}) = \frac{1}{2} y^{*2}_{i,r}$, where $y^*_{i,r}$ denote the equilibrium effort level.

the above example. When it is the aggregate effort of friends that affects one's utility, the position in the network affects one's equilibrium effort. When the deviation from the social norm is costly, all individuals want to conform to the effort of their reference group, which is the same for all of them when they are *ex ante* identical.

### 2.5 Discussion

Let us discuss the differences between local-average and local-aggregate models and what kinds of mechanisms they imply in more detail. For that, we will start with Whyte's (1955) study of the Italian North End of Boston in the late 1930's. Whyte studied the behavior of a street-corner gang, especially that of their leader Doc. Whyte wondered why Doc, a highly intelligent and curious individual, was not upwardly mobile and, instead, dropped out of school. Whyte was puzzled by Doc's behavior because school would have been easy for Doc given his exceptional ability and intelligence. Whyte concluded that Doc did not seek extra education out of loyalty to his group, whom he would be abandoning were he to advance beyond them educationally. The behavior of Doc is in accordance with the local-average model where it is costly to deviate from the group's social norm. Even if Doc is much more intelligent than the members of his gang, it would be too costly for him to acquire a higher level of education since this would mean interacting less with his friends or even abandoning them. Contrary to a model with no social interactions, where educational costs are mainly tuition fees, lost wages, etc., here it is the cost of lost contacts with one's friends that is crucial. Now, if Doc had preferences according to the local-aggregate model, he would have acted differently. His decision to seek extra education would have been driven by his formidable ability and the sum of his friends' educational level, which is going to be quite high as Doc, a leader, has many friends. What is crucial, however, is that there would not be a cost from deviating from his friends' decisions and he would certainly have decided to pursue education, despite the lower average education level of his peers.

This means that the policy implications of the two models are quite different. In the local-average model, the only way to affect individuals' behavior and thus their outcomes is to change the *social norm* of the group. In other words, one needs to affect most people in the group for the policy to be effective. As a result, *group-based policies*, for example, a school-based or a region-based policy, should be implemented in the context of this model. On the other hand, in the local-aggregate model, because of *social multiplier* effects, one can target only one individual and still have positive effects because she, in turn, will affect her peers. In that case, individual-based policy could be implemented.

Akerlof (1997) discussed Eugene Lang's famous offer to give a college scholarship to *every student* at the sixth grade class in Harlem. Of the 51 students who remained in the New York area, 40 were considered likely to go to college six years later. Akerlof (1997) explained the success of this policy by the fact that it affected *all* students not some of them. As Akerlof put it: "The experiment was successful because the students formed a cohesive group in which each member received reinforcement from others who, like themselves, were on the academic track toward graduation from high school". In the language of the local-average model, this policy worked well because it changes the norm's group by affecting all its members. After the policy experiment, graduating and going to college was not anymore considered as "bad" but as the social norm of the group, i.e. what should be done. In the context of the local-aggregate model, one does not need to undertake such a costly policy. It suffices to give a college scholarship to some students who, by increasing their performance, will increase the total effort of peer reference group of their friends, who will, in turn, affect the total effort of their own friends, etc.

# 3 Identification of the Econometric Network Model

## 3.1 Econometric network models

The specification of the econometric model follows the equilibrium best-reply function of the network game so that it has a clear microfoundation. Let the *ex ante* heterogeneity $\pi_{i,r}$ of individual $i$ in network $r$ be

$$\pi_{i,r} = x'_{i,r}\beta + \sum_{j=1}^{n_r} g^*_{ij,r} x'_{j,r}\gamma + \eta_r + \epsilon_{i,r},$$

where $x_{i,r}$ is an $p$-dimensional vector of exogenous variables, $\epsilon_{i,r}$ is an i.i.d. innovation with zero mean and finite variance $\sigma^2$, and $\beta, \gamma, \eta_r$ are corresponding parameters. From the best-reply function (2), the general econometric network model is

$$y_{i,r} = \phi_1 \sum_{j=1}^{n_r} g_{ij,r} y_{j,r} + \phi_2 \sum_{j=1}^{n_r} g^*_{ij,r} y_{j,r} + x'_{i,r}\beta + \sum_{j=1}^{n_r} g^*_{ij,r} x'_{j,r}\gamma + \eta_r + \epsilon_{i,r}, \tag{8}$$

for $i = 1, \cdots, n_r$ and $r = 1, \cdots, \bar{r}$. Let $Y_r = (y_{1,r}, \cdots, y_{n_r,r})'$, $X_r = (x_{1,r}, \cdots, x_{n_r,r})'$, and $\epsilon_r = (\epsilon_{1,r}, \cdots, \epsilon_{n_r,r})'$. Then, (8) can be written in matrix form as

$$Y_r = \phi_1 G_r Y_r + \phi_2 G_r^* Y_r + X_r \beta + G_r^* X_r \gamma + \eta_r l_{n_r} + \epsilon_r.$$

Let $\mathrm{diag}\{A_j\}$ denote a "generalized" block diagonal matrix in which the diagonal blocks are $m_j \times n_j$ matrices $A_j$'s. For a data set with $\bar{r}$ groups, let $Y = (Y_1', \cdots, Y_{\bar{r}}')'$, $X = (X_1', \cdots, X_{\bar{r}}')'$, $\eta = (\eta_1, \cdots, \eta_{\bar{r}})'$, $\epsilon = (\epsilon_1', \cdots, \epsilon_{\bar{r}}')'$, $G = \mathrm{diag}\{G_r\}_{r=1}^{\bar{r}}$, $G^* = \mathrm{diag}\{G_r^*\}_{r=1}^{\bar{r}}$ and $L = \mathrm{diag}\{l_{n_r}\}_{r=1}^{\bar{r}}$. The general econometric model can be written as

$$Y = \phi_1 GY + \phi_2 G^* Y + X\beta + G^* X\gamma + L\eta + \epsilon. \tag{9}$$

The econometric network model (9) incorporates *the endogenous effect*, captured by $\phi_1$ and $\phi_2$, where an individual's choice/outcome may depend on those of her friends, *the contextual effect*, captured by the coefficient $\gamma$, where an individual's choice/outcome may depend on the exogenous characteristics of her friends, and *the network fixed effect*, captured by the network-specific parameter $\eta_r$, where individuals in the same network may behave similarly as they have similar unobserved individual characteristics or they face a similar institutional environment (see Manski, 1993).[11] Furthermore, we distinguish between *the aggregate endogenous effect*, captured by the coefficient $\phi_1$, and *the average endogenous effect*, captured by the coefficient $\phi_2$, as they originate from different economic models with totally different equilibrium implications.

The network-specific parameters $\eta$ is allowed to depend on $G$, $G^*$ and $X$ as in a fixed effect panel data model. To avoid the incidental parameter problem when the number of groups $\bar{r}$ is large, we eliminate the term $L\eta$ using the deviation from group mean projector $J = \mathrm{diag}\{J_r\}_{r=1}^{\bar{r}}$, 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 a fixed effect panel data model. As $JL = 0$, the transformed network model is

$$JY = \phi_1 JGY + \phi_2 JG^* Y + JX\beta + JG^* X\gamma + J\epsilon. \tag{10}$$

[11] Network fixed effects can be motivated by a two-step network formation model where agents self-select into different networks in a first step and, then, in a second step, link formation takes place within networks based on observable individual characteristics only. Therefore, the network fixed effect serves as a (partial) remedy for the selection bias that originates from the possible sorting of individuals with similar unobserved characteristics into a network.

It is well known that endogenous and contextual effects cannot be separately identified in a linear-in-means model due to the *reflection problem*, first formulated by Manski (1993). The reflection problem arises because, in a linear-in-means model, individuals are affected by all individuals belonging to their group and by nobody outside the group, and thus the simultaneity in behavior of individuals in the same group introduces a perfect collinearity between the endogenous effect and the contextual effect. For the network model, the reference group usually varies across individuals and the identification of various social interaction effects thus becomes possible. Let $Z = [GY, G^*Y, X, G^*X]$ denote the matrices of regressors in the network model. Let $Q$ denote the IV matrix.[12] Then, the network model (9) is identified if the following condition is satisfied.

**Identification Condition** $\lim_{n\to\infty} \frac{1}{n} Q'\mathrm{E}(JZ)$ is a finite matrix with full column rank.

This identification condition implies the rank condition that $\mathrm{E}(JZ)$ has full column rank and that the column rank of $Q$ is at least as high as that of $\mathrm{E}(JZ)$, for large enough $n$. In the rest of this section, we provide sufficient conditions for this identification condition. Based on the identification result, the transformed model can be estimated by generalizing the 2SLS and GMM methods in Liu and Lee (2010).[13]

### 3.2 Identification of the local-average model

First, we consider the identification of (9) under the exclusion restriction $\phi_1 = 0$. In this case, the general network model reduces to the local-average model

$$Y = \phi_2 G^*Y + X\beta + G^*X\gamma + L\eta + \epsilon, \tag{11}$$

with the reduced-form equation

$$Y = (I - \phi_2 G^*)^{-1}(X\beta + G^*X\gamma + L\eta + \epsilon). \tag{12}$$

[12]For the local-average model, a possible IV matrix is $Q = J[X, G^*X, G^{*2}X]$, where $G^{*2}X$ represents the characteristics of friends' friends. For the local-aggregate model, a possible IV matrix is $Q = J[X, G^*X, G^{*2}X, GL]$, where $GL$ represents the number of friends.

[13]See Section 4.2 and Appendix B at the end of the paper for more details.

As $J(I-\phi_2 G^*)^{-1}L = 0$ and $(I-\phi_2 G^*)^{-1} = \sum_{j=0}^{\infty}(\phi_2 G^*)^j$, it follows from (12) that

$$\mathrm{E}(JG^*Y) = JG^*X\beta + JG^{*2}(I-\phi_2 G^*)^{-1}X(\phi_2\beta+\gamma) = JG^*X\beta + (JG^{*2}X + \phi_2 JG^{*3}X + \cdots)(\phi_2\beta+\gamma). \tag{13}$$

To illustrate the challenges for the identification of the local-average model, we consider the two following cases:

(i) $\phi_2\beta + \gamma = 0$. In this case, $\mathrm{E}(JG^*Y) = JG^*X\beta$. The model cannot be identified because, for $Z = [G^*Y, X, G^*X]$, $\mathrm{E}(JZ) = [\mathrm{E}(JG^*Y), JX, JG^*X]$ does not have full column rank. From the perspective of IV estimation, $\mathrm{E}(JG^*Y) = JG^*X\beta$ implies that the only informative IV for the endogenous effect is $JG^*X$, which is also a regressor in the transformed model, the contextual effect. The perfect collinearity between the mean of the endogenous effect and the contextual effect makes the model unidentified. A special case of $\phi_2\beta + \gamma = 0$ is $\beta = \gamma = 0$. In this case, $\mathrm{E}(JG^*Y) = 0$ and, thus, the model cannot be identified as there is no relevant IV for the endogenous effect.

(ii) $\phi_2\beta + \gamma \neq 0$. In this case, identification of the local-average model may be possible as, according to (13), $[JG^{*2}X, JG^{*3}X, \cdots]$ can be used IVs for the endogenous effect. Note, in a natural network, if individuals $i, j$ are friends and $j, k$ are friends, it does not necessarily imply that $i, k$ are also friends. The *intransitivity in social connections* provides an exclusion restriction such that the characteristics of the friends' friends $G^{*2}X$ may not be perfectly correlated with own characteristics $X$ and the characteristics of the friends $G^*X$. Thus, one can use IVs like $JG^{*2}X$ to identify endogenous and contextual effects. Based on this important observation, Bramoullé et al. (2009) have shown that if the matrices $I, G^*, G^{*2}, G^{*3}$ are linearly independent, social effects are identified. From the perspective of a spatial autoregressive (SAR) model, Lee et al. (2010) have shown that model (11) can be identified if $[JX, JG^*X, JG^{*2}X]$ has full column rank.

### 3.3 Identification of the local-aggregate model

On the other hand, under the exclusion restriction $\phi_2 = 0$, the general network model (9) reduces to the local-aggregate model

$$Y = \phi_1 GY + X\beta + G^*X\gamma + L\eta + \epsilon. \tag{14}$$

For the $r$-th network, the reduced-form equation of the local-aggregate model is

$$Y_r = (I_{n_r} - \phi_1 G_r)^{-1}(X_r\beta + G_r^* X_r \gamma + \eta_r l_{n_r} + \epsilon_r),$$

which implies

$$\mathrm{E}(J_r G_r Y_r) = J_r G_r (I_{n_r} - \phi_1 G_r)^{-1}(X_r\beta + G_r^* X_r \gamma) + \eta_r J_r G_r (I_{n_r} - \phi_1 G_r)^{-1} l_{n_r}. \tag{15}$$

Note, the term $G_r\left(I_{n_r} - \phi_1 G_r\right)^{-1} l_{n_r}$ is the Bonacich measure of centrality (Ballester et al. 2006; Bonacich, 1987). When row sums of $G_r$ are not constant, the term $J_r G_r\left(I_{n_r} - \phi_1 G_r\right)^{-1} l_{n_r} \neq 0$ and therefore can be used as an additional IV for model identification.

As the identification condition requires $\mathrm{E}(JZ)$, where $Z = [GY, X, G^*X]$ for the local-aggregate model, to have full column rank, the following proposition gives a sufficient condition for the rank condition. Henceforth, let $c$ (possibly with subscripts) denote a constant scalar that may take different values for different uses.

Let also

$$\Lambda_1 = \begin{bmatrix} \gamma c_1 & 1 & -\phi_1 c_1 \\ \beta + \gamma c_2 & -\phi_1 & -\phi_1 c_2 \\ \gamma c_3 & 0 & 1 - \phi_1 c_3 \\ \eta_r & 0 & 0 \end{bmatrix} \tag{16}$$

and

$$\Lambda_2 = \begin{bmatrix} -\gamma c_1 & 1 & \phi_1 c_1 \\ \beta - \gamma c_2 & -\phi_1 & \phi_1 c_2 \\ -\gamma c_3 & -1 & \phi_1 c_3 + 1 \\ \gamma - \beta - \gamma c_4 & \phi_1 & \phi_1 c_4 - \phi_1 \\ -\gamma c_5 & 0 & \phi_1 c_5 - 1 \end{bmatrix}. \tag{17}$$

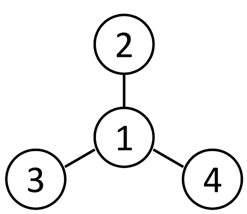
We have the following result:

**Proposition 2** *For the local-aggregate network model (14), let* $\mathrm{E}(JZ) = [\mathrm{E}(JGY), JX, JG^*X]$.

- *When $G_r$ has non-constant row sums for some network $r$, $\mathrm{E}(JZ)$ has full column rank if:* (i) *$I_{n_r}, G_r, G_r^*, G_r G_r^*$ are linearly independent and $|\beta| + |\gamma| + |\eta_r| \neq 0$; or* (ii) *$G_r G_r^* = c_1 I_{n_r} + c_2 G_r + c_3 G_r^*$ and $\Lambda_1$ given by (16) has full rank.*

- *When $G_r$ has constant row sums such that $g_{i,r} = g_r$ for all $r$,* $\mathrm{E}(JZ)$ *has full column rank if:* (iii) *$I, G, G^*, GG^*, G^{*2}, GG^{*2}$ are linearly independent and $|\beta|+|\gamma| \neq 0$;* (iv) *$I, G, G^*, GG^*, G^{*2}$ are linearly independent, $GG^{*2} = c_1 I + c_2 G + c_3 G^* + c_4 GG^* + c_5 G^{*2}$, and $\Lambda_2$ given by (17) has full rank; or* (v) *$g_r = g$ for all $r$, $I, G^*, G^{*2}, G^{*3}$ are linearly independent, and $\phi_1 \beta g + \gamma \neq 0$.*

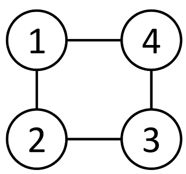
In many cases, the identification conditions for the *local-aggregate* model given in Proposition 2 are weaker than those for the *local-average* model given in Bramoullé et al. (2009). Figure 2 gives an example where identification is possible for the local-aggregate model but fails for the local-average model. Consider a data set where each network is represented by the graph in Figure 2 (a star-shaped network). The adjacency matrix $G$ is a block-diagonal matrix with diagonal blocks being $G_r$ in Figure 2. For the row-normalized adjacency matrix $G^*$, it is easy to see that $G^{*3} = G^*$. Therefore, it follows from Proposition 5 of Bramoullé et al. (2009) that the local-average model (11) is not identified. On the other hand, as $G_r$ in Figure 2 has non-constant row sums and $I_{n_r}, G_r, G_r^*, G_r G_r^*$ are linearly independent, it follows from our Proposition 2(i) that the local-aggregate model (14) can be identified for this network. From the perspective of IVs, as $J(G^*)^{j+2}X = (-1)^{j+1} JG^*X$ for $j = 0, 1, 2, \cdots$, the IVs based on (13) $[JG^{*2}X, JG^{*3}X, \cdots]$ are all perfectly correlated with the contextual effect regressor $JG^*X$. Therefore, the local-average model cannot be identified. On the other hand, for the local-aggregate model, as $G_r$ does not have constant row sums, one can use $J_r G_r l_{n_r}$, the leading-order term of $J_r G_r (I_{n_r} - \phi_1 G_r)^{-1} l_{n_r}$ in (15), as an IV to identify endogenous and contextual effects.



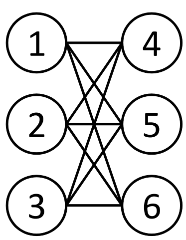

$$G_r = \begin{bmatrix} 0 & 1 & 1 & 1 \\ 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \end{bmatrix} \quad \text{and} \quad G_r^* = \begin{bmatrix} 0 & 1/3 & 1/3 & 1/3 \\ 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \end{bmatrix}.$$

Figure 2: An example where the local-aggregate model can be identified by Proposition 2(i).

Figure 3 provides another example where the local-average model cannot be identified while the local-aggregate model can. Consider a data set with two types of networks. The first type of network is represented by the graph on the top of Figure 3 (a regular network or a circle). The second type of network is represented by the graph on the bottom of Figure 3 (a bi-partite network). For these two networks, the adjacency matrix $G$ is a block-diagonal matrix with diagonal blocks being either $G_1$ or $G_2$ given in Figure 3. For the row normalized adjacency matrix $G^*$, it is easy to see that $G^{*3} = G^*$. Therefore, it follows from Proposition 5 of Bramoullé et al. (2009) that the local-average model (11) is not identified. On the other hand, as the two different types of networks have different row sums, $I, G, G^*, GG^*, G^{*2}$ are linearly independent and $GG^{*2} = G$. Therefore, the local-aggregate model (14) can be identified by our Proposition 2(iv). From the perspective of IVs, the local-average model cannot be identified because the IVs based on (13) $[JG^{*2}X, JG^{*3}X, \cdots]$ are all perfectly correlated with the contextual effect regressor $JG^*X$. On the other hand, for the local-aggregate model, as $JGX$ and $JGG^*X$ are linearly independent of the contextual effect regressor $JG^*X$, one can use $JGX$ and $JGG^*X$, the leading-order terms of $JG(I - \phi_1 G)^{-1}(X\beta + G^*X\gamma)$ in (15), as IVs to identify endogenous and contextual effects.




$$G_1 = \begin{bmatrix} 0 & 1 & 0 & 1 \\ 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \\ 1 & 0 & 1 & 0 \end{bmatrix} \qquad G_1^* = \begin{bmatrix} 0 & 1/2 & 0 & 1/2 \\ 1/2 & 0 & 1/2 & 0 \\ 0 & 1/2 & 0 & 1/2 \\ 1/2 & 0 & 1/2 & 0 \end{bmatrix},$$




$$G_2 = \begin{bmatrix} 0 & 0 & 0 & 1 & 1 & 1 \\ 0 & 0 & 0 & 1 & 1 & 1 \\ 0 & 0 & 0 & 1 & 1 & 1 \\ 1 & 1 & 1 & 0 & 0 & 0 \\ 1 & 1 & 1 & 0 & 0 & 0 \\ 1 & 1 & 1 & 0 & 0 & 0 \end{bmatrix} \qquad G_2^* = \begin{bmatrix} 0 & 0 & 0 & 1/3 & 1/3 & 1/3 \\ 0 & 0 & 0 & 1/3 & 1/3 & 1/3 \\ 0 & 0 & 0 & 1/3 & 1/3 & 1/3 \\ 1/3 & 1/3 & 1/3 & 0 & 0 & 0 \\ 1/3 & 1/3 & 1/3 & 0 & 0 & 0 \\ 1/3 & 1/3 & 1/3 & 0 & 0 & 0 \end{bmatrix}.$$

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

### 3.4 Identification of the general network model

Let us now consider the identification of the general network model. For the $r$-th network, the reduced-form equation is

$$Y_r = (I_{n_r} - \phi_1 G_r - \phi_2 G_r^*)^{-1}(X_r\beta + G_r^* X_r \gamma + \eta_r l_{n_r} + \epsilon_r),$$

which implies

$$\mathrm{E}(J_rG_rY_r) = J_rG_r(I_{n_r} - \phi_1 G_r - \phi_2 G_r^*)^{-1}(X_r\beta + G_r^* X_r\gamma) + J_rG_r(I_{n_r} - \phi_1 G_r - \phi_2 G_r^*)^{-1} l_{n_r}\eta_r, \quad (18)$$

and

$$\mathrm{E}(J_rG_r^*Y_r) = J_rG_r^*(I_{n_r} - \phi_1 G_r - \phi_2 G_r^*)^{-1}(X_r\beta + G_r^* X_r\gamma) + J_rG_r^*(I_{n_r} - \phi_1 G_r - \phi_2 G_r^*)^{-1} l_{n_r}\eta_r. \quad (19)$$

First, we consider the case when all networks have constant row sums such that $g_{i,r} = g_r$ for all $r$. If, furthermore, $g_r = g$ for all $r$, i.e., all network have the same degrees (row sums), then $G = gG^*$ and the general network model cannot be identified as the local-aggregate endogenous variable $GY$ and local-average endogenous variable $G^*Y$ are perfectly correlated. If there are at least two networks in the data that have different degrees so that $G$ and $G^*$ are linearly independent, then the general network model can be identified through the following proposition.

**Proposition 3** *Suppose $G_r$ has constant row sums such that $g_{i,r} = g_r$ for all $r$. $\mathrm{E}(JZ)$ of the general network model (9) has full column rank if $I, G, G^*, GG^*, G^{*2}, GG^{*2}, G^{*3}$ are linearly independent and $\phi_1\beta \neq 0$ or $\gamma + \phi_2\beta \neq 0$.*

Intuitively, the identification result given by Proposition 3 relies on the variation of degrees $g_r$ across networks. Note that if $\beta = \gamma = 0$, i.e., there is no relevant exogenous covariate in the model, then the identification condition given by Proposition 3 does not hold.

When the row sums of $G_r$ are not constant for some network $r$ and $\eta_r \neq 0$, then it follows from (18) that $J_rG_r(I_{n_r} - \phi_1 G_r - \phi_2 G_r^*)^{-1} l_{n_r}$ can be used as an IV for the local-aggregate endogenous variable $J_rG_rY_r$. Furthermore, if $\phi_1 \neq 0$, then it follows from (19) that $J_rG_r^*(I_{n_r} - \phi_1 G_r - \phi_2 G_r^*)^{-1} l_{n_r}$ can be used as an IV for the local-average endogenous variable $J_rG_r^*Y_r$. In this case, the general network model may still be identifiable even if there is no relevant exogenous covariate in the model such that $\beta = \gamma = 0$.

## 4 Local Aggregate or Local Average? A Specification Test

From the above discussion, we can see that the local-average and local-aggregate models have different equilibrium implications and identification conditions. In this section, we propose a statistical model selection test to detect which behavioral mechanism better represents the data. As local-average and local-aggregate models are special cases of the general network model (9), one could simply test for individual significance of $\phi_1$ or $\phi_2$ in (9). Here, we consider a more general specification test for network models, which can be applied to, but not limited to, the test for local-average and local-aggregate effects.

In standard linear regression models, the J test is used to compare non-nested model specifications (Davidson and MacKinnon, 1981). The idea of the J test is as follows. If a given model contains the correct set of regressors, then including the fitted values of an alternative model (or of a fixed number of competing models) into the null model should provide no significant improvement.

Kelejian (2008) extends the J test to a spatial framework. He shows that the test could, but need not, relate solely to the specification of the spatial weighting matrix. Importantly, since the J test relies on whether the prediction based on an alternative model significantly increases the explanatory power of the null model, it is important to use all the available information in the alternative model. However, Kelejian (2008) does not use the information in an efficient way to determine the predictions (Kelejian and Piras, 2011). Here, the J test of Kelejian (2008) and Kelejian and Piras (2011) is implemented using the spatial 2SLS estimation procedure. Our paper's contributions in this regard can be summarized as follows.

(1) We generalize the J test in Kelejian and Piras (2011) to a network model with network fixed effects. Our source of identification of the augmented model is the variation in the row sums of the adjacency matrix $G$.

(2) We first consider the 2SLS estimation of the augmented model to implement the J test. Besides the IVs proposed by Kelejian and Prucha (1998), we consider additional IVs based on the vector of degrees $G_r l_{n_r}$ (the number of friends) to improve identification and estimation efficiency. The number of such IVs is the same as the number of networks in the data. If the number of IVs is large relative to our sample size, the 2SLS estimator could be asymptotically biased (Liu and Lee, 2010). Hence, we propose a bias-correction procedure to eliminate the leading order many-IV bias.

(3) When the IVs are weak and thus the J test based on the 2SLS estimator is not reliable, we propose a GMM estimator to implement the J test. The GMM estimator uses additional quadratic moment conditions, which are especially helpful when the IVs are weak.

### 4.1 J test for model selection

The local-aggregate and local-average models (14) and (11) can be written more compactly as:

$$H_1 \quad : \quad Y = \phi_1 GY + X^* \delta_1 + L\eta_1 + \epsilon_1, \tag{20}$$

$$H_2 \quad : \quad Y = \phi_2 G^* Y + X^* \delta_2 + L\eta_2 + \epsilon_2, \tag{21}$$

where $X^* = [X, G^* X]$, and $\delta_1, \delta_2$ are corresponding vector of coefficients.

#### 4.1.1 The test of model $H_1$ against model $H_2$

To test against the model specification $H_2$, one can estimate the following augmented model of $H_1$,

$$Y = \alpha_1 Y_{H_2} + \phi_1 GY + X^* \delta_1 + L\eta_1 + \epsilon_1, \tag{22}$$

where $Y_{H_2}$ is a predictor of $Y$ under $H_2$ such that $Y_{H_2} = \phi_2 G^* Y + X^* \delta_2 + L\eta_2$ (see Kelejian and Prucha, 2007; Kelejian and Piras, 2011). Thus, a test of the null model (20) against the alternative one (21) would be in terms of the hypotheses: $H_0 : \alpha_1 = 0$ against $H_a : \alpha_1 \neq 0$.

Substitution of the predictor $Y_{H_2}$ into (22) gives

$$\begin{aligned} Y &= \alpha_1(\phi_2 G^* Y + X^* \delta_2) + \phi_1 GY + X^* \delta_1 + L(\eta_1 + \alpha_1 \eta_2) + \epsilon_1 \\ &= Z_1^* \vartheta_1 + L(\eta_1 + \alpha_1 \eta_2) + \epsilon_1, \end{aligned} \tag{23}$$

where $Z_1^* = [\phi_2 G^* Y + X^* \delta_2, GY, X^*]$ and $\vartheta_1 = (\alpha_1, \phi_1, \delta_1')'$. The within transformation of (23) with the deviation from group mean projector $J$ gives

$$JY = JZ_1^* \vartheta_1 + J\epsilon_1. \tag{24}$$

The proposed J test can be implemented by the following two steps:

(1) Estimate model $H_2$ by the quasi-maximum-likelihood (QML) method of Lee et al. (2010). Let the preliminary QML estimators of $\phi_2$ and $\delta_2$ be denoted by $\tilde{\phi}_2$ and $\tilde{\delta}_2$.

(2) Estimate the feasible counterpart of model (24)

$$JY = J\tilde{Z}_1^* \vartheta_1 + J\epsilon_1, \tag{25}$$

where $\tilde{Z}_1^* = [\tilde{\phi}_2 G^* Y + X^* \tilde{\delta}_2, GY, X^*]$, by the 2SLS or GMM method described in Section 4.2. If the estimated $\alpha_1$ is insignificant, then this is evidence against model $H_2$.

#### 4.1.2 The test of model $H_2$ against model $H_1$

The test of model $H_2$ against model $H_1$ can be carried out in a similar manner. Consider the following augmented model of $H_2$,

$$H_2 : Y = \alpha_2 Y_{H_1} + \phi_2 G^* Y + X^* \delta_2 + L\eta_2 + \epsilon_2, \tag{26}$$

where $Y_{H_1}$ is a predictor of $Y$ under $H_1$ such that $Y_{H_1} = \phi_1 GY + X^* \delta_1 + L\eta_1$.[14] Thus, the test of the null model (21) against the alternative (20) would be in terms of the hypotheses $H_0 : \alpha_2 = 0$ against $H_a : \alpha_2 \neq 0$. The within transformation of (26) with the projector $J$ gives

$$JY = JZ_2^* \vartheta_2 + J\epsilon_2. \tag{27}$$

where $Z_2^* = [\phi_1 GY + X^* \delta_1, G^* Y, X^*]$ and $\vartheta_2 = (\alpha_2, \phi_2, \delta_2)'$.

The proposed J test can be implemented by the following two steps:

(1) Estimate model $H_1$ by the 2SLS with IVs $J[X, G^* X, GX]$. Let the preliminary 2SLS estimators of $\phi_1$ and $\delta_1$ be denoted by $\tilde{\phi}_1$ and $\tilde{\delta}_1$.

(2) Estimate the feasible counterpart of model (27)

$$JY = J\tilde{Z}_2^* \vartheta_2 + J\epsilon_2, \tag{28}$$

where $\tilde{Z}_2^* = [(\tilde{\phi}_1 GY + X^* \tilde{\delta}_1), G^* Y, X^*]$, by the 2SLS or GMM method described in Section 4.2. If the estimated $\alpha_2$ is significant, then that is evidence against model $H_1$.

### 4.2 The 2SLS and GMM estimators

For the estimation of the general network model (10) and the augmented models (25) or (28) in the second step of the J test, we consider the following estimators by generalizing the 2SLS and GMM methods in Liu and Lee (2010):

[14] An alternative predictor of $Y$ under $H_1$ is $Y_{H_1} = (I - \phi_1 G)^{-1} X^* \delta_1 + (I - \phi_1 G)^{-1} L\eta_1$ (see Kelejian, 2008 and Kelejian and Piras, 2011). However, this predictor brings some additional complication to the estimation of the augmented model

$$Y = \alpha_2 (I - \phi_1 G)^{-1} X^* \delta_1 + \phi_2 G^* Y + X^* \delta_2 + \alpha_2 (I - \phi_1 G)^{-1} L\eta_1 + L\eta_2 + \epsilon_2$$

as the term $\alpha_2 (I - \phi_1 G)^{-1} L\eta_1$ cannot be eliminated by the projector $J$.

(a) "2SLS": a 2SLS estimator with the IV matrix $Q_1 = J[X, G^*X, GX, G^{*2}X]$.

(b) "BC2SLS": a bias-corrected 2SLS estimator with the IV matrix $Q_2 = J[X, G^*X, GX, G^{*2}X, GL]$. The additional IVs $JGL$ correspond to the information on different positions of group members measured by Bonacich (1987) centrality. The additional IVs improve asymptotic efficiency of the estimator and help achieve identification when the "conventional" IVs $Q_1$ are weak. Note that, the matrix $JGL$ has $\bar{r}$ columns, where $\bar{r}$ is the number of networks in the data. Therefore, if there are many networks (e.g. in the empirical study, there are 490 networks in our data), the 2SLS estimator with the IV matrix $Q_2$ may have an asymptotic bias, which is known as the many-instrument bias.[15] The "BC2SLS" estimator corrects the many-instrument bias by an estimated leading-order bias term.

The 2SLS estimators are based on moment conditions that are *linear* in the model coefficients. However, when the IVs are weak, the inference based on the 2SLS estimation may be unreliable. Lee (2007a) has suggested to generalize the 2SLS method to a comprehensive GMM framework with additional *quadratic* moment conditions based on the covariance structure of the reduced form equation to improve identification and estimation efficiency. The added quadratic moment conditions are especially helpful when the IVs are weak. In this paper, we consider the following GMM estimators for the estimation of the empirical model:

(c) "GMM": an optimal GMM estimator using linear moment conditions with $Q_1$ and quadratic moment conditions.

(d) "BCGMM": a bias-corrected optimal GMM estimator using linear moment conditions with $Q_2$ and the same quadratic moment conditions as in "GMM". Similar to the corresponding 2SLS estimator, the additional IVs in $Q_2$ may introduce a many-instrument bias into the GMM estimator. The "BCGMM" estimator corrects the many-instrument bias by an estimated leading-order bias term.

The details of the 2SLS and GMM methods, including the explicit form of the quadratic moment condition, are given in Appendix B.

[15] This is less of a concern in the data used in this paper, as the number of groups are small relative to the sample size.

## 5 Empirical Application

### 5.1 Data description

As an illustration of the proposed econometric methodology, we present an empirical analysis of study effort (i.e. education) and sport activities among teenagers in the United States. Our analysis is made possible by the use of a unique database on friendship networks from the National Longitudinal Survey of Adolescent Health (AddHealth). The AddHealth database has been designed to study the impact of the social environment (i.e. friends, family, neighborhood and school) on adolescents' behavior in the United States by collecting data on students in grades 7-12 from a nationally representative sample of roughly 130 private and public schools in years 1994-95. Every pupil attending the sampled schools on the interview day is asked to compile a questionnaire (in-school data) containing questions on respondents' demographic and behavioral characteristics, education, family background and friendship. This sample contains information on roughly 90,000 students.[16]

From a network perspective, the most interesting aspect of the AddHealth data is the information on friendships. Indeed, the friendship information is based upon actual friend nominations. Students were asked to identify their best friends from a school roster (up to five males and five females).[17] Knowing exactly who nominates whom in a network, we exploit the directed nature of the nominations data.[18] We focus on choices made and we denote a link from $i$ to $j$ as $g_{ij,r} = 1$ if $i$ has nominated $j$ as her friend in network $r$, and $g_{ij,r} = 0$, otherwise.[19] By matching the identification numbers of the friendship nominations to respondents' identification numbers, one can obtain information on the characteristics of nominated friends. More importantly, one can reconstruct the whole geometric structure of the friendship networks. For each school, we thus obtain all the networks of (best) friends.[20,21]

[16] A subset of adolescents selected from the rosters of the sampled schools, about 20,000 individuals, is then asked to compile a longer questionnaire containing more sensitive individual and household information (in-home and parental data). Those subjects of the subset are interviewed again in 1995–96 (wave II), in 2001–2 (wave III), and again in 2007-2008 (wave IV). For the purpose of our analysis, we only focus on wave I in-school data.

[17] The limit in the number of nominations is not binding (even by gender). Less than 1% of the students in our sample show a list of ten best friends.

[18] We also consider the undirected nature of the friendship relationships in Section 5.3.

[19] As highlighted by Wasserman and Faust (1994), centrality indices for directional relationships generally focus on choices made.

[20] Note that, when an individual $i$ identifies a best friend $j$ who does not belong to the same school, the database does not include $j$ in the network of $i$; it provides no information about $j$. Fortunately, in the large majority of cases (more than 93%), best friends tend to be in the same school and thus are systematically included in the network.

[21] We construct networks as *network components*. Network components are maximally connected networks, which satisfy the two following conditions. First, two agents in a network component are either directly linked or are indirectly linked through a sequence of agents (this is the requirement of *connectedness*). Second, two agents in different network components cannot be connected through any such sequence (this is *maximality*). A school usually

We exploit this unique data set to understand the impact of peer pressure on individual behavior for two different outcomes: (i) study effort[22] and (ii) sport activities. The "study effort" corresponds to the response to the question: "In general, how hard do you try to do your school work well?", coded using an ordinal scale as 0 (I never try at all), 1 (I don't try very hard), 2 (I try hard enough/but not as hard as I could), and 3 (I try very had to do my best). Involvement in sport activities is measured by the response to the question: "How many times in a normal week to you work, play, or exercise hard enough to make you sweat and breathe heavily?", coded as 0 (never), 1 (1 or 2 times), 2 (3 to 5 times), 3 (6 or 7 times), and 4 (more than 7 times). Definitions and summary statistics of the variables used in our empirical analysis can be found in Table C.1 in Appendix C.

After removing students with missing or inadequate information, isolated students and pairs (i.e. network with only two students), the sample consists of 63,325 students distributed over 215 networks, with network size ranging from 3 to 1,833. Because the strength of peer effect may vary with network sizes (see Calvó-Armengol et al., 2009) and the endogeneity of friendships may be more severe in large and sparse networks, we focus our analysis on networks with a size of between 3 and 300 students.[23] Our selected sample consists of 8,607 students distributed over 133 networks, with network sizes ranging from 3 to 299.[24] The mean and the standard deviation of network size are 64.71 and 89.90. Furthermore, in our sample, the average number of friends of a student is 3.60 with the standard deviation 2.58.

### 5.2 Estimation results

Tables 1 and 2 report the estimation results for the general econometric network model (9) using alternative estimators for "study effort" (Table 1) and "sport activities" (Table 2). We consider the 2SLS and GMM estimators described in Section 4.2 and the bias-corrected 2SLS and GMM the estimators that use additional IVs based on Bonacich centrality (i.e. BC2SLS and BCGMM described in Section 4.2). Tables 1 and 2 also report the first stage F test statistic and the over-identifying restrictions (OIR) test p-value for the 2SLS estimator. The IVs described in Section 4.2 are only valid if the network adjacency matrix $G$ is exogenous conditional on control variables $X$ and network fixed effects. We find that the p-values of the OIR test are larger than conventional

contains more than one network.

[22] Since the outcome of our theoretical model, $y_{i,r}$, is effort in some activity, it makes sense to use "study effort" as a measure of education rather than the grade of the student, even if the latter is positively correlated to the former.

[23] Our results, however, do not depend crucially on these network size thresholds. They remain qualitatively unchanged when slightly moving the network size window.

[24] From the summary statistics given in Table C.1, we can see that the selected sample is representative of the original data.

significance level for both activities, which provides evidence that $G$ can be considered as exogenous in this empirical analysis.

[Insert Tables 1 and 2 here]

Do peer effects matter? Which model is more adequate for each activity? These questions can be answered from the statistical significance of $\phi_1$ and $\phi_2$ of the general network model or from the p-values of the J test reported in Tables 1 and 2.

Table 1 reveals that, for study effort, the endogenous peer effect is mostly captured by a *social-conformity effect* rather than a social-multiplier effect. In other words, the local-average model is the most appropriate model for education as measured by the "study effort" of each student. This is in line with the findings in Lin (2010) and Boucher et al. (2014). On the other hand, for sport activities, we find in Table 2 that both social-conformity and social-multiplier effects contribute to the endogenous peer effect. In terms of magnitude, the BCGMM estimates suggest that, for a student with 10 friends, if the sport activity index of these 10 students increases by 1 point, then my sporting activity index will increase by about 0.1 points because of my desire to conform to the social norm of the group and by about 0.07 points because of the social multiplier effect.[25]

[Insert Table 3 here]

Our results are interesting and new. First, they show that different forms of social interactions may drive peer effects in different activities. Second, they show that even for the same activity there might be different mechanisms of peer effects at work. In this respect, our findings suggest notes of caveat in the empirical analysis of peer effects. Peer effects are a complex phenomenon and their assessment should be considered with caution. If more than one mechanism is driving social interactions, then neglecting one of them can produce biased inferential results. In Table 3, we report the empirical results obtained when *separately* estimating the local-aggregate and the local-average model. Comparing Tables 1, 2 and 3, it appears clearly that the local-aggregate peer effect is overstated if the local-average effect is ignored and vice versa. The bias is even more severe, when we model "study effort" using the local-aggregate model since the J test suggests that the

[25] As the maximum number of friends a student can nominate is 10, $g_r^{\max} \leq 10$. The constraint for the uniqueness of the equilibrium $g_r^{\max}\phi_1 + \phi_2 < 1$ holds for all specifications and estimators considered.

local-average model is more appropriate for this activity.

### 5.3 Robustness checks

As a robustness check, we first consider the estimation of the general network model without including network fixed effects. The estimation results are reported in Table 4. The network fixed effect captures the correlated effect at the network level. Without network fixed effects, the estimates of the endogenous peer effects reported in Table 4 are upwards biased relative to those reported in Tables 1 and 2. However, the qualitative results remain unchanged.

[Insert Table 4 here]

Our identification and estimation strategies depend on the correct specification of network links. In particular, our identification strategy hinges upon nonlinearities in friendship, i.e. on the presence of intransitive triads. In this section, we test the robustness of our results with respect to misspecification of network topology. So far, we have measured peer groups as precisely as possible by exploiting the direction of the nomination data. However, friendship relationships are reciprocal in nature, and even if a best friend of a given student does not nominate this student as her best friend, one may think that social interactions take place. Under this circumstance, there can be some "unobserved" network link that, if considered, would change the network topology and break some intransitivities in network links. Therefore, in this section, we repeat our analysis by considering *undirected networks*, i.e. we assume that a link exists between two friends only if both students have nominated each other, that is $g_{ij,r} = g_{ji,r} = 1$. Table 5 reports the main results for undirected networks. The qualitative results remain unchanged.

[Insert Table 5 here]

### 5.4 Policy implications

We would like to discuss the different policy implications of the local-aggregate and local-average network models. Indeed, we believe that it is important to be able to disentangle between different behavioral peer-effect models because they have different policy implications. We base our discussion on our empirical results in Section 5.2 where the activity considered is education (i.e. "study effort"). This is a context where peer effects matter and where policy interventions are crucial.

Indeed, education is clearly an important topic and effective policies are difficult to implement.[26] There has been some debate in the United States of giving incentives to teachers. It is, however, difficult to determine which incentive to give to teachers in order to improve teacher quality. If the local aggregate model is at work among teachers, then we would need to have a teacher-based incentive policy since teachers will influence each other while, if it is the local average model, then one should implement a school-based incentive policy because this will be the only way to change the social norm of working hard among teachers.

If we now consider the students themselves, then the two models will be useful for policy implications. If the local-aggregate model is important in explaining students' education outcomes (Calvó-Armengol et al., 2009), then any individual-based policy (for example, *vouchers*) would be efficient. If, on the contrary, as shown in the present paper, we believe that the local-average model is more important, then we should change the social norm in the school or the classroom and try to implement the idea that it is "cool" to work hard at school.[27] For example, in Section 2.5, we discussed Eugene Lang's famous offer to give a college scholarship to *every student* at the sixth grade class in Harlem. This policy worked well because it changed the norm's group by affecting all its members.

An example of a policy that has tried to change the social norm of students in terms of education is the *charter-school* policy. The charter schools are very good in screening teachers and at selecting the best ones. In particular, the "No Excuses policy" (Angrist et al., 2010, 2012) is a highly standardized and widely replicated charter model that features a long school day, an extended school year, selective teacher hiring, strict behavior norms, and emphasizes traditional reading and math skills. The main objective is to change the social norms of disadvantage kids by being very strict on discipline. This is a typical policy that is in accordance with the local-average model since its aim is to change the social norm of students in terms of education. Angrist et al. (2012) focus on special needs students that may be underserved. Their results show average achievement gains of 0.36 standard deviations in math and 0.12 standard deviations in reading for each year spent at a charter school called: Knowledge is Power Program (KIPP) Lynn, with the largest gains coming from the Limited English Proficient (LEP), Special Education (SPED), and low-achievement groups.

[26] It has to be clear that the policy implications discussed in this section are heuristic and not derived formally.

[27] This is related to the "acting white" literature where it is argued that African American students in poor areas may be ambivalent about studying hard in school because this may be regarded as "acting white" and adopting mainstream identities (Fordham and Ogbu, 1986; Delpit, 1995; Ainsworth-Darnell and Downey, 1998; Austen-Smith and Fryer, 2005; Battu et al., 2007; Fryer and Torelli, 2010; Battu and Zenou, 2010; Bisin et al., 2011; de Martí and Zenou, 2014).

They show that the average reading gains were driven almost entirely by SPED and LEP students, whose reading scores rose by roughly 0.35 standard deviations for each year spent at KIPP Lynn.[28]

The local-average model can also help us design an adequate policy in terms of tracking at school (Betts, 2011). Should we "track" students in a way that separates high achievers from low achievers or should we mix them? If we believe that the local-average model matters, then the answer is that we should separate high achievers from low achievers but then have an exogenous intervention on the low achievers in order to change their social norms. A way to do so is to send them to a charter school as the Angrist et al. (2012) study suggests. However, if the local aggregate mechanism of peer effects prevails, then classes should be heterogenous with respect to students' test scores, with the high performing students distributed among the classes. Under this scenario, high achievers will have a positive impact on low achievers but will not be able to change the social norm of the low achievers.

To sum-up, an effective policy for the *local-average* model would be to change people's perceptions of "normal" behavior (i.e. their social norm) so that a *school-based policy* should be implemented while, for the *local-aggregate* model, this would not be necessary and an *individual-based policy* should instead be implemented.

## 6 Concluding remarks

Identifying the nature of peer effects is a topic as important for policy purposes as difficult to study empirically. While a variety of mechanisms have been put forward in the theoretical literature, the econometrics of networks is lagging behind. This paper develops a unified econometric framework to estimate two types of social interaction (peer group) effects based on a given network structure. We provide a micro foundation by exploring different types of utility functions and illustrate the methodology using an application to education and sport activities. Our results show that different forms of social interactions may drive peer effects in different outcomes. Furthermore, they show that even for the same outcome there might be different mechanisms of peer effects at work. In this respect, our findings suggest some notes of caveat in the empirical analysis of peer effects. Peer effects are a complex phenomenon and their assessment should be considered with caution. If more than one mechanism is driving social interactions, then neglecting one of them can produce biased

[28] See also Curto and Fryer (2014) who study the SEED schools, which combine a "No Excuses" charter model with a 5-day-a-week *boarding* program. These are America's only urban public boarding schools for the poor. Using admission lotteries, they show that attending a SEED school increases achievement by 0.211 standard deviation in reading and 0.229 standard deviation in math per year.

inferential results.

## References


[1] Ainsworth-Darnell, J.W. and D.B. Downey (1998), "Assessing the oppositional culture explanation for racial/ethnic differences in school performance," *American Sociological Review* 63, 536-553.

[2] Akerlof, G.A. (1980), "A theory of social custom of which unemployment may be one consequence," *Quarterly Journal of Economics* 94, 749-775.

[3] Akerlof, G.A. (1997), "Social distance and social decisions," *Econometrica* 65, 1005-1027.

[4] Angrist, J.D., Dynarski, S.M., Kane, T.J., Pathak, P.A. and C.R. Walters (2010), "Inputs and impacts in charter schools: KIPP Lynn," *American Economic Review Papers and Proceedings* 100, 239-243.

[5] Angrist, J.D., Dynarski, S.M., Kane, T.J., Pathak, P.A. and C.R. Walters (2012), "Who benefits from KIPP?" *Journal of Policy Analysis and Management* 31, 837-860.

[6] Austen-Smith D. and R.D. Fryer (2005), "An economic analysis of 'acting white'," *Quarterly Journal of Economics* 120 551-583.

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

[8] Ballester, C., Calvó-Armengol, A. and Y. Zenou (2010), "Delinquent networks," *Journal of the European Economic Association* 8, 34-61.

[9] Battu, H., M. Mwale, and Y. Zenou (2007), "Oppositional identities and the labor market," *Journal of Population Economics* 20, 643-667.

[10] Battu, H. and Y. Zenou (2010), "Oppositional identities and employment for ethnic minorities. Evidence for England," *Economic Journal* 120, F52-F71.

[11] Berman, E. (2000), "Sect, subsidy, and sacrifice: An economist's view of ultra-orthodox Jews," *Quarterly Journal of Economics* 115, 905-953.

[12] Bernheim, B.D. (1994), "A theory of conformity," *Journal of Political Economy* 102, 841-877.

[13] Betts, J.R. (2011), "The economics of tracking in education," In: E.A. Hanushek, S.J. Machin and L. Woessmann (Eds.), *Handbook of the Economics of Education*, Volume 3, Amsterdam: Elsevier Science, pp. 341-381.

[14] Bisin, A., Patacchini, E., Verdier, T. and Y. Zenou (2011), "Formation and persistence of oppositional identities," *European Economic Review* 55, 1046-1071.

[15] Bonacich, P. (1987), "Power and centrality: A family of measures," *American Journal of Sociology* 92, 1170-1182.

[16] Boucher, V., Bramoullé, Y., Djebbari, H. and B. Fortin (2014), "Do peers affect student achievement? Evidence from Canada using group size variation," *Journal of Applied Econometrics* 29, 91-109.

[17] Bramoullé, Y., Djebbari, H. and B. Fortin (2009), "Identification of peer effects through social networks," *Journal of Econometrics* 150, 41-55.

[18] Bramoullé, Y. and R.E. Kranton (2007), "Public goods in networks," *Journal of Economic Theory* 135, 478-494.

[19] Calvó-Armengol, A., Patacchini, E. and Y. Zenou (2005), "Peer effects and social networks in education and crime," CEPR Discussion Paper No. 5244.

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

[21] Clark, A.E. and Y. Loheac (2007), "It wasn't me, it was them! Social influence in risky behavior by adolescents," *Journal of Health Economics* 26, 763-784.

[22] Clark, A.E. and A.J. Oswald (1998), "Comparison-concave utility and following behaviour in social and economic settings," *Journal of Public Economics* 70, 133-155.

[23] Chao, J, C., Hausman, J., Newey, W.K., Swanson, N.R. and T.Woutersen (2010), "Testing overidentifying restrictions with many instruments and heteroskedasticity," University of Maryland Working Paper.

[24] Curto, V.E. and R.G. Fryer Jr. (2014), "The potential of urban boarding schools for the poor: Evidence from SEED," *Journal of Labor Economics* 32, 65-93.

[25] Cvetković, D. and P. Rowlinson (1990), "The largest eigenvalue of a graph: A survey," *Linear and Multilinear Algebra* 28, 3-33.

[26] Davidson R. and J.G. MacKinnon (1981), "Several tests for model specification in the presence of alternative hypotheses," *Econometrica* 49, 781-793.

[27] Davidson R. and J.G. MacKinnon (1993), *Estimation and inference in Econometrics*, New York: Oxford University Press.

[28] de Martí, J. and Y. Zenou (2014), "Identity and social distance in friendship formation," Unpublished manuscript, Stockholm University.

[29] Delpit, L. (1995), *Other People's Children: Cultural Conflict in the Classroom*, New York: The Free Press.

[30] Durlauf, S. (2004), "Neighborhood effects," In: J.V. Henderson and J-F. Thisse (Eds.), *Handbook of Regional and Urban Economics Vol. 4*, Amsterdam: Elsevier Science.

[31] Fershtman, C., and Y. Weiss (1998), "Social rewards, externalities and stable preferences," *Journal of Public Economics* 70, 53-73.

[32] Fordham, S. and Ogbu, J.U. (1986), "Black student'school success: Coping with the burden of 'acting white'," *The Urban Review* 18, 176-206.

[33] Fryer, R.G. and P. Torelli (2010), "An empirical analysis of 'acting white'," J*ournal of Public Economics* 94, 380–396.

[34] Galeotti A., Goyal S., Jackson M.O., Vega-Redondo F. and L. Yariv (2009), "Network games," *Review of Economic Studies* 77, 218-244.

[35] Ghiglino, C. and S. Goyal (2010), "Keeping up with the neighbors: Social interaction in a market economy," *Journal of the European Economic Association* 8, 90-119.

[36] Glaeser, E.L. and J. Scheinkman (2001), "Measuring social interactions," In: S. Durlauf and H.P. Young (Eds.), *Social Dynamics*, Cambridge: MIT Press.

[37] Glaeser, E.L. and J. Scheinkman (2003), "Non-market interactions", In: M. Dewatripont, L. P. Hansen and S. Turnovsky (Eds.), *Advances in Economics and Econometrics: Theory and Applications*, Cambridge: Cambridge University Press.

[38] Goyal, S. (2007), *Connections: An Introduction to the Economics of Networks*, Princeton: Princeton University Press.

[39] Horn, R. and C. Johnson (1990), *Matrix Analysis*, Cambridge: Cambridge University Press.

[40] Iannaccone, L. R. (1998), "Introduction to the economics of religion," *Journal of Economic Literature* 36, 1465-1496.

[41] Ioannides, Y.M. and D.L. Loury (2004), "Job information networks, neighborhood effects, and inequality," *Journal of Economic Literature* 42, 1056-1093.

[42] Ioannides, Y.M. (2012), *From Neighborhoods to Nations: The Economics of Social Interactions*, Princeton: Princeton University Press.

[43] Jackson, M.O. (2008), *Social and Economic Networks*, Princeton: Princeton University Press.

[44] Jackson, M.O. and Y. Zenou (2013), "Games on networks," In: P. Young and S. Zamir (Eds.), *Handbook of Game Theory, Vol. 4*, Amsterdam: Elsevier, forthcoming.

[45] Kandel, E. and, E.P. Lazear (1992), "Peer pressure and partnerships," *Journal of Political Economy* 100, 801-817.

[46] Kelejian, H.H. (2008), "A spatial J-test for model specification against a single or a set of nonnested alternatives," *Letters in Spatial and Resources Sciences* 1, 3-11.

[47] Kelejian, H.H. and I. Prucha (1998), "A generalized spatial two-stage least squares procedure for estimating a spatial autoregressive model with autoregressive disturbances," *Journal of Real Estate Finance and Economics* 17, 99-121

[48] Kelejian, H.H. and I.R. Prucha (2007), "The relative efficiencies of various predictors in spatial econometric models containing spatial lags," *Regional Science and Urban Economics* 37, 363-374.

[49] Kelejian, H.H. and G. Piras (2011), "An extension of Kelejian's J-test for non-nested spatial models," *Regional Science and Urban Economics* 41, 281-292.

[50] Lee, L.F. (2002), "Consistency and efficiency of least squares estimation for mixed regressive, spatial autoregressive models," *Econometric Theory* 18, 252-277.

[51] Lee, L.F. (2003), "Best spatial two-stage least squares estimators for a spatial autoregressive model with autoregressive disturbances," *Econometric Reviews* 22, 307-335.

[52] Lee, L.F. (2007a), "GMM and 2SLS estimation of mixed regressive, spatial autoregressive models," *Journal of Econometrics* 137, 489-514

[53] Lee, L.F. (2007b), "Identification and estimation of econometric models with group interactions, contextual factors and fixed effects," *Journal of Econometrics* 140, 333-374.

[54] Lee, L.F. and X. Liu (2010), "Efficient GMM estimation of high order spatial autoregressive models with autoregressive disturbances," *Econometric Theory* 26, 187-230.

[55] Lee, L.F., Liu, X. and X. Lin (2010), "Specification and estimation of social interaction models with network structures," *Econometrics Journal* 13, 145-176.

[56] Lin, X. (2010), "Identifying peer effects in student academic achievement by a spatial autoregressive model with group unobservables," *Journal of Labor Economics* 28, 825-860.

[57] Liu, X. and L.F. Lee (2010), "GMM estimation of social interaction models with centrality," *Journal of Econometrics* 159, 99-115.

[58] Liu, X., Patacchini, E., Zenou, Y. and L-F. Lee (2012), "Criminal networks: Who is the key player?" Unpublished manuscript, Stockholm University, available at: http://people.su.se/~yvze0888/.

[59] Manski, C.F. (1993), "Identification of endogenous effects: The reflection problem," *Review of Economic Studies* 60, 531-542.

[60] Meyer, C.D. (2000), *Matrix Analysis and Applied Linear Algebra*, Philadelphia, PA: SIAM.

[61] Moffitt, R. (2001), "Policy interventions low-level equilibria, and social interactions," In: S. Durlauf and P. Young (Eds.), *Social Dynamics*, Cambridge, MA: MIT Press, pp. 45-82.

[62] Patacchini, E. and Y. Zenou (2008), "The strength of weak ties in crime," *European Economic Review* 52, 209-236.

[63] Patacchini, E. and Y. Zenou (2012), "Juvenile delinquency and conformism," *Journal of Law, Economic, and Organization* 28, 1-31.

[64] Stock, J.H and M. Yogo (2005), "Testing for weak instruments in linear IV regression," In: D.W.K. Andrews and J.H. Stock (Eds.), *Identification and Inference for Econometric Models: Essays in Honor of Thomas Rothenberg*, Cambridge: Cambridge University Press, pp. 80-108.

[65] Stock, J.H, Wright J.H. and M. Yogo (2002), "A survey of weak instruments and weak identification in generalized method of moments," *Journal of Business and Economic Statistics* 2, 518-529.

[66] Sutherland, E.H. (1947), *Principles of Criminology*, fourth edition, Chicago: J.B. Lippincott.

[67] Wasserman, S. and K. Faust (1994), *Social Network Analysis. Methods and Applications*, Cambridge: Cambridge University Press.

[68] Whyte, W.F. (1955), *Street Corner Society*, Second Edition, Chicago: University of Chicago Press.

# APPENDICES

## A Proofs

**Proof of Proposition 1.** Since $G_r^*$ is the row-normalized $G_r$, we have $G_r = R_r G_r^*$, where $R_r = \text{diag}\{g_{i,r}\}_{i=1}^{n_r}$ and $g_{i,r} = \sum_{j=1}^{n_r} g_{ij,r}$. Let $||\cdot||_\infty$ denote the row-sum matrix norm. As $||G_r^*||_\infty = 1$, we have

$$||\phi_1 G_r + \phi_2 G_r^*||_\infty = ||\phi_1 R_r G_r^* + \phi_2 G^*||_\infty \leq ||\phi_1 R_r + \phi_2 I_{n_r}||_\infty = g_r^{\max}\phi_1 + \phi_2 < 1.$$

Hence, $I_{n_r} - \phi_1 G_r - \phi_2 G_r^*$ is invertible (see Horn and Johnson, 1990) and the desired result follows. ■

**Proof of Proposition 2.** See Liu et al. (2012). ■

**Proof of Proposition 3.** Suppose $G_r$ has constant row sums such that $g_{i,r} = g_r$ for all $r$. Then, $G_r = g_r G_r^*$ and $G = RG^*$ where $R = \text{diag}\{g_r I_{n_r}\}_{r=1}^{\bar{r}}$. $\text{E}(JZ_3) = [\text{E}(JGY), \text{E}(JG^*Y), JX, JG^*X]$ has full column rank if

$$[\text{E}(JGY)d_1 + \text{E}(JG^*Y)d_2 + JXd_3 + JG^*Xd_4] = 0 \tag{29}$$

implies $d_1 = d_2 = d_3 = d_4 = 0$. As $G = RG^*$, we have $JG(I - \phi_1 G - \phi_2 G^*)^{-1}L = JG^*(I - \phi_1 G - \phi_2 G^*)^{-1}L = 0$, $G(I - \phi_1 G - \phi_2 G^*)^{-1} = (I - \phi_1 G - \phi_2 G^*)^{-1}G$ and $G^*(I - \phi_1 G - \phi_2 G^*)^{-1} = (I - \phi_1 G - \phi_2 G^*)^{-1}G^*$. Then, it follows from (18) and (19) that

$$\text{E}(JGY) = J(I - \phi_1 G - \phi_2 G^*)^{-1}G(X\beta + G^*X\gamma), \tag{30}$$

$$\text{E}(JG^*Y) = J(I - \phi_1 G - \phi_2 G^*)^{-1}G^*(X\beta + G^*X\gamma). \tag{31}$$

Plugging (30) and (31) into (29) gives

$$J(I-\phi_1 G-\phi_2 G^*)^{-1}[Xd_3+GX(\beta d_1-\phi_1 d_3)+G^*X(\beta d_2-\phi_2 d_3+d_4)+GG^*X(\gamma d_1-\phi_1 d_4)+G^{*2}X(\gamma d_2-\phi_2 d_4)] = 0,$$

which implies

$$Xd_3+GX(\beta d_1-\phi_1 d_3)+G^*X(\beta d_2-\phi_2 d_3+d_4)+GG^*X(\gamma d_1-\phi_1 d_4)+G^{*2}X(\gamma d_2-\phi_2 d_4) = \rho L, \tag{32}$$

for a constant scalar $\rho$. Premultiplying (32) by $G^*$ gives

$$G^*Xd_3+GG^*X(\beta d_1-\phi_1 d_3)+G^{*2}X(\beta d_2-\phi_2 d_3+d_4)+GG^{*2}X(\gamma d_1-\phi_1 d_4)+G^{*3}X(\gamma d_2-\phi_2 d_4)=\rho L. \tag{33}$$

From (32) and (33), when $I, G, G^*, GG^*, G^{*2}, GG^{*2}, G^{*3}$ are linearly independent, then $d_1 = d_2 = d_3 = d_4 = 0$ if $\phi_1\beta \neq 0$ or $\gamma + \phi_2\beta \neq 0$. ■

# B 2SLS and GMM Estimation

We consider 2SLS and GMM estimators for the estimation of an empirical general network model, and for the estimation of augmented models in the J test. This appendix presents the derivation and asymptotic properties of the estimators.

For any $n\times n$ matrix $A = [a_{ij}]$, let $\mathrm{vec}_D(A) = (a_{11}, \cdots, a_{nn})'$, $A^s = A+A'$, $A^t = A-\mathrm{tr}(A)J/\mathrm{tr}(J)$, and $A^-$ denote a generalized inverse of a square matrix $A$. For a parameter $\theta$, let $\theta_0$ denote the true parameter value in the data generating process. Let $\mu_3$ and $\mu_4$ denote, respectively, the third and fourth moments of the error term.

## B.1 Estimation of the general network model

### B.1.1 2SLS estimation

Let $M_0 = (I - \phi_{10}G - \phi_{20}G^*)^{-1}$. From the reduced form equation, $\mathrm{E}(Y) = M_0(X^*\delta_0 + L\eta)$.[29] For $Z = [GY, G^*Y, X^*]$, the ideal IV matrix for the explanatory variables $JZ$ in (10) is given by

$$f = \mathrm{E}(JZ) = J[G\mathrm{E}(Y), G^*\mathrm{E}(Y), X^*]. \tag{34}$$

However, this IV matrix is infeasible as it involves unknown parameters. Note that $f$ can be considered as a linear combination of the IVs in $Q_\infty = J[GM_0X^*, GM_0L, G^*M_0X^*, G^*M_0L, X^*]$. As $L$ has $\bar{r}$ columns, the number of IVs in $Q_\infty$ increases as the number of groups $\bar{r}$ increases. Furthermore, if $|\phi_{10}\max_i(\sum_j g_{ij})| + |\phi_{20}| < 1$,[30] we have $M_0 = (I - \phi_{10}G - \phi_{20}G^*)^{-1} = \sum_{j=0}^{\infty}(\phi_{10}G + \phi_{20}G^*)^j$. Hence, $M_0$ in $Q_\infty$ can be approximated by a linear combination of $[I, G, G^*, G^2, GG^*, G^*G, G^{*2}, \cdots]$.

[29] For simplicity, we assume $G$ and $X$ are nonstochastic. If $G$ and $X$ are stochastic, then the following results can be considered as conditional on $G$ and $X$.

[30] The model represents an equilibrium so $I - \phi_{10}G - \phi_{20}G^*$ is assumed to be invertible. In Proposition 1, we showed that a sufficient condition for the invertibility assumption is: $|\phi_{10}d^{\max}| + |\phi_{20}| < 1$, where $d^{\max} \equiv \max_i g_i$ is the highest degree in network $G$. On the other hand, a sufficient condition for the the invertibility of $I - \phi_{10}G$ for the local aggregate model is $|\phi_{10}|d^{\max} < 1$ and a sufficient condition for the the invertibility of $I - \phi_{20}G^*$ for the local average model is $|\phi_{20}| < 1$. Both of them are weaker than the invertibility condition of the hybrid model.

To achieve asymptotic efficiency, we assume the number of IVs increases with the sample size so that the ideal IV matrix $f$ can be approximated by a feasible IV matrix $Q_K$ with an approximation error diminishing to zero. That is, for an $n \times K$ IV matrix $Q_K$ premultiplied by $J$, there exists some conformable matrix $\xi_K$ such that $||f - Q_K\xi_K||_\infty \to 0$ as $n, K \to \infty$. Let $P_K = Q_K(Q_K'Q_K)^- Q_K'$, the 2SLS estimator is given by $\hat{\theta}_{2sls} = (Z'P_KZ)^{-1}Z'P_KY$.

Let $\theta_0 = (\phi_{10}, \phi_{20}, \delta_0')'$. If $K/n \to 0$, then it follows by a similar argument as in Liu and Lee (2010) that $\sqrt{n}(\hat{\theta}_{2sls} - \theta_0 - b_{2sls}) \xrightarrow{d} N(0, \sigma^2\bar{H}^{-1})$, where $\bar{H} = \lim_{n\to\infty} \frac{1}{n} f'f$ and $b_{2sls} = \sigma^2(Z'P_KZ)^{-1}[\mathrm{tr}(P_KGM_0), \mathrm{tr}(P_KG^*M_0), 0_{1\times 2m}]' = O_p(K/n)$. The 2SLS estimator has an asymptotic bias term due to the large number of IVs. When $K^2/n \to 0$, the leading order bias term $\sqrt{n}b_{2sls}$ converges to zero and the proposed 2SLS estimator is efficient as the variance matrix $\sigma^2\bar{H}^{-1}$ attains the efficiency lower bound for the class of IV estimators.

To correct for the many-instrument bias in the 2SLS estimator, one can estimate the leading order bias term and adjust the 2SLS estimator by the estimated leading-order bias $\tilde{b}_{2sls}$. With $\sqrt{n}$-consistent initial estimates $\check{\sigma}^2, \check{\phi}_1, \check{\phi}_2$, the bias-corrected 2SLS (BC2SLS) is given by $\hat{\theta}_{c2sls} = \hat{\theta}_{2sls} - \tilde{b}_{2sls}$, where $\tilde{b}_{2sls} = \check{\sigma}^2(Z'P_KZ)^{-1}[\mathrm{tr}(P_KGM), \mathrm{tr}(P_KG^*M), 0_{1\times 2m}]'$ and $M = (I - \check{\phi}_1G - \check{\phi}_2G^*)^{-1}$. The BC2SLS is efficient when $K/n \to 0$.

### B.1.2 GMM estimation

The 2SLS estimator can be generalized to the GMM with additional quadratic moment equations. Let $\epsilon(\theta) = J(Y - Z\theta)$. The IV moment conditions $Q_K'\epsilon(\theta) = 0$ are linear in $\epsilon$ at $\theta_0$. As $\mathrm{E}(\epsilon'U_1\epsilon)$ $=\mathrm{E}(\epsilon'U_2\epsilon) = 0$ for $U_1 = (JGM_0J)^t$ and $U_2 = (JG^*M_0J)^t$, the quadratic moment conditions for estimation are given by $[U_1\epsilon(\theta), U_2\epsilon(\theta)]'\epsilon(\theta) = 0$. The proposed quadratic moment conditions can be shown to be optimal (in terms of efficiency of the GMM estimator) under normality (see Lee and Liu, 2010). The vector of linear and quadratic empirical moments for the GMM estimation is given by $g(\theta) = [Q_K, U_1\epsilon(\theta), U_2\epsilon(\theta)]'\epsilon(\theta)$.

In order for inference based on the following asymptotic results to be robust, we do not impose the normality assumption for the following analysis. The variance matrix of $g(\theta_0)$ is given by

$$\Omega = \mathrm{Var}[g(\theta_0)] = \begin{pmatrix} \sigma^2 Q_K'Q_K & \mu_3 Q_K'\omega \\ \mu_3\omega'Q_K & (\mu_4 - 3\sigma^4)\omega'\omega + \sigma^4\Upsilon \end{pmatrix},$$

where $\omega = [\mathrm{vec}_D(U_1), \mathrm{vec}_D(U_2)]$ and $\Upsilon = \frac{1}{2}[\mathrm{vec}(U_1^s), \mathrm{vec}(U_2^s)]'[\mathrm{vec}(U_1^s), \mathrm{vec}(U_2^s)]$. By the generalized

Schwarz inequality, the optimal GMM estimator is given by

$$\hat{\theta}_{gmm} = \arg\min g'(\theta)\Omega^{-1}g(\theta). \tag{35}$$

Let $B^{-1} = (\mu_4 - 3\sigma^4)\omega'\omega + \sigma^4\Upsilon - \frac{\mu_3^2}{\sigma^2}\omega' P_K \omega$,

$$D = -\sigma^2 \begin{pmatrix} \mathrm{tr}(U_1^s G M_0) & \mathrm{tr}(U_1^s G^* M_0) & 0_{1\times 2m} \\ \mathrm{tr}(U_2^s G M_0) & \mathrm{tr}(U_2^s G^* M_0) & 0_{1\times 2m} \end{pmatrix},$$

$\bar{D} = D - \frac{\mu_3}{\sigma^2}\omega' f$, and $\check{D} = D - \frac{\mu_3}{\sigma^2}\omega' P_K Z$. When $K^{3/2}/n \to 0$, the optimal GMM estimator[31] has the asymptotic distribution

$$\sqrt{n}(\hat{\theta}_{gmm} - \theta_0 - b_{gmm}) \xrightarrow{d} N(0, (\sigma^{-2}\bar{H} + \lim_{n\to\infty} \frac{1}{n}\bar{D}' B \bar{D})^{-1}), \tag{36}$$

where $b_{gmm} = (\sigma^{-2} Z' P_K Z + \check{D}' B \check{D})^{-1}[\mathrm{tr}(P_K G M_0), \mathrm{tr}(P_K G^* M_0), 0_{1\times 2m}]' = O(K/n)$.

As the asymptotic bias $\sqrt{n} b_{gmm}$ is $O(K/\sqrt{n})$, the asymptotic distribution of the GMM estimator $\hat{\theta}_{gmm}$ will be centered at $\theta_0$ only if $K^2/n \to 0$. With a consistently estimated leading order bias $\tilde{b}_{gmm}$, the bias-corrected GMM (BCGMM) estimator $\hat{\theta}_{cgmm} = \hat{\theta}_{gmm} - \tilde{b}_{gmm}$ has a proper centered asymptotic normal distribution as given in (36) if $K^{3/2}/n \to 0$.

The asymptotic variance matrix of the many-IV GMM estimator can be compared with that of the many-IV 2SLS estimator. As $\bar{D}' B \bar{D}$ is nonnegative definite, the asymptotic variance of the many-IV GMM estimator is smaller relative to that of the 2SLS estimator. Thus, the many-IV GMM estimator with additional quadratic moments improves efficiency upon the 2SLS estimator.

### B.2 Estimation of augmented models in the J test

In this subsection, we focus on the estimation of the augmented model in the test of model $H_1$ against model $H_2$. The estimator for the test of model $H_2$ against model $H_1$ can be derived in a similar manner.

[31] The weighting matrices for quadratic moments $U_1, U_2$ and the optimal weighting matrix for the objective function $\Omega^{-1}$ involves unknown parameters $\phi_1, \phi_2, \sigma_0^2, \mu_3$ and $\mu_4$. With consistent preliminary estimators of those unknown parameters, the feasible optimal GMM estimator can be shown to have the same asymptotic distribution given by (36).

### B.2.1 2SLS estimation of the augmented model

First, we consider the 2SLS estimator of the augmented model (24). Let $M_{10} = (I - \alpha_{10}\phi_{20}G^* - \phi_{10}G)^{-1}$. The ideal IV matrix for $JZ_1^*$ in (24) is given by $f_1 = \mathrm{E}(JZ_1^*) = J[\phi_{20}G^*\mathrm{E}(Y) + X^*\delta_{20}, G\mathrm{E}(Y), X^*]$, where $\mathrm{E}(Y) = M_{10}[X^*(\alpha_{10}\delta_{20} + \delta_{10}) + L(\eta_1 + \alpha_{10}\eta_2)]$. The ideal IV matrix $f_1$ is infeasible as it involves unknown parameters. We note that $f_1$ can be considered as a linear combination of the IVs in $Q_\infty = J[G^*M_{10}X^*, G^*M_{10}L, GM_{10}X^*, GM_{10}L, X^*]$. Furthermore, under some regularity conditions, $M_{10} = (I - \alpha_{10}\phi_{20}G^* - \phi_{10}G)^{-1} = \sum_{j=0}^{\infty}(\alpha_{10}\phi_{20}G^* + \phi_{10}G)^j$. Hence, $M_{10}$ in $Q_\infty$ can be approximated by polynomials of $I$, $G$ and $G^*$.

To achieve asymptotic efficiency, we consider an $n \times K$ feasible submatrix of $Q_\infty$, denoted by $Q_K$, such that the ideal IV matrix $f_1$ can be approximated by a linear combination of $Q_K$ with an approximation error diminishing to zero as the number of IVs $K$ increases. Let $P_K = Q_K(Q_K'Q_K)^-Q_K'$ and $\tilde{Z}_1^* = [(\tilde{\phi}_2G^*Y + X^*\tilde{\delta}_2), GY, X^*]$, where $\tilde{\phi}_2, \tilde{\delta}_2$ are $\sqrt{n}$-consistent preliminary estimates. The 2SLS estimator considered is $\hat{\vartheta}_{1,tsls} = (\tilde{Z}_1^{*\prime}P_K\tilde{Z}_1^*)^{-1}\tilde{Z}_1^{*\prime}P_KY$.

Under the null hypothesis, it follows by a similar argument as in Liu and Lee (2010) that if $K/n \to 0$ then $\sqrt{n}(\hat{\vartheta}_{1,tsls} - \vartheta_{10} - b_{1,tsls}) \xrightarrow{d} N(0, \sigma_1^2\bar{H}_1^{-1})$, where $\bar{H}_1 = \lim_{n\to\infty}\frac{1}{n}f_1'f_1$ and $b_{1,tsls} = \sigma_1^2(\tilde{Z}_1^{*\prime}P_K\tilde{Z}_1^*)^{-1}[\phi_{20}\mathrm{tr}(P_KG^*M_{10}), \mathrm{tr}(P_KGM_{10}), 0_{1\times 2m}]'$. The term $b_{1,tsls}$ is a bias due to the presence of many IVs. We can adjust for the many-IV bias by considering the BC2SLS estimator $\hat{\vartheta}_{1,ctsls} = \hat{\vartheta}_{1,tsls} - \tilde{b}_{1,tsls}$, where $\tilde{b}_{1,tsls}$ is a consistent estimator of $b_{1,tsls}$. If $K/n \to 0$ then $\sqrt{n}(\hat{\vartheta}_{1,ctsls} - \vartheta_{10}) \xrightarrow{d} N(0, \sigma_1^2\bar{H}_1^{-1})$.

### B.2.2 GMM estimation of the augmented model

The GMM estimator uses both linear moment conditions $Q_K'\epsilon_1(\vartheta_1) = 0$ and quadratic ones

$$[U_1\epsilon_1(\vartheta_1), U_2\epsilon_1(\vartheta_1)]'\epsilon_1(\vartheta_1) = 0,$$

where $U_1 = (JG^*M_{10}J)^t$, $U_2 = (JGM_{10}J)^t$, and $\epsilon_1(\vartheta_1) = J(Y - \tilde{Z}_1^*\vartheta_1)$. The vector of linear and quadratic empirical moment functions for the GMM estimation is given by $g_1(\vartheta_1) = [Q_K, U_1\epsilon_1(\vartheta_1), U_2\epsilon_1(\vartheta_1)]'\epsilon_1(\vartheta_1)$. By the generalized Schwarz inequality, the optimal GMM estimator is given by $\hat{\vartheta}_{1,gmm} = \arg\min g_1'(\vartheta_1)\Omega^{-1}g_1(\vartheta_1)$, where

$$\Omega = \begin{pmatrix} \sigma_1^2 Q_K' Q_K & \mu_3 Q_K'\omega \\ \mu_3\omega' Q_K & (\mu_4 - 3\sigma_1^4)\omega'\omega + \sigma_1^4\Upsilon \end{pmatrix},$$

$\omega = [\mathrm{vec}_D(U_1), \mathrm{vec}_D(U_2)]$ and $\Upsilon = \frac{1}{2}[\mathrm{vec}(U_1^s), \mathrm{vec}(U_2^s)]'[\mathrm{vec}(U_1^s), \mathrm{vec}(U_2^s)]$.

Let $B_1^{-1} = (\mu_4 - 3\sigma_1^4)\omega'\omega + \sigma_1^4\Upsilon - \frac{\mu_3^2}{\sigma_1^2}\omega' P_K\omega$,

$$D_1 = -\sigma_1^2 \begin{pmatrix} \phi_{20}\mathrm{tr}(U_1^s G^* M_{10}) & \mathrm{tr}(U_1^s G M_{10}) & 0_{1\times 2m} \\ \phi_{20}\mathrm{tr}(U_2^s G^* M_{10}) & \mathrm{tr}(U_2^s G M_{10}) & 0_{1\times 2m} \end{pmatrix},$$

$\bar{D}_1 = D_1 - \frac{\mu_3}{\sigma_1^2}\omega' f_1$, and $\check{D}_1 = D_1 - \frac{\mu_3}{\sigma_1^2}\omega' P_K \tilde{Z}_1^*$. Under the null hypothesis, if $K^{3/2}/n \to 0$, the optimal GMM estimator[32] has the asymptotic distribution

$$\sqrt{n}(\hat{\vartheta}_{1,gmm} - \vartheta_{10} - b_{1,gmm}) \xrightarrow{d} N(0, (\sigma_1^{-2}\bar{H}_1 + \lim_{n\to\infty}\frac{1}{n}\bar{D}_1' B_1 \bar{D}_1)^{-1}), \tag{37}$$

where $b_{1,gmm} = (\sigma_1^{-2}\tilde{Z}_1^{*\prime} P_K \tilde{Z}_1^* + \check{D}_1' B_1 \check{D}_1)^{-1}[\phi_{20}\mathrm{tr}(P_K G^* M_{10}), \mathrm{tr}(P_K G M_{10}), 0_{1\times 2m}]' = O(K/n)$.

With a consistently estimated leading order bias $\tilde{b}_{1,gmm}$, it follows by a similar argument as in Liu and Lee (2010) that, if $K^{3/2}/n \to 0$, the BCGMM estimator $\hat{\vartheta}_{1,cgmm} = \hat{\vartheta}_{1,gmm} - \tilde{b}_{1,gmm}$ has a proper centered asymptotic normal distribution as given in (37).

[32] With consistent preliminary estimates of the unknown parameters in $U_1, U_2, \Omega$, the feasible optimal GMM estimator can be shown to have the same asymptotic distribution given by (37).

# Appendix C: Data appendix

Table C.1: Data Summary

| Variable | Definition | Sample with networks 3-1833 ($n = 63325$) | | Sample with networks 3-300 ($n = 8607$) | |
|---|---|---|---|---|---|
| | | Mean | SD | Mean | SD |
| study effort | In the text | 2.24 | 0.67 | 2.38 | 0.66 |
| sport activity | In the text | 2.28 | 1.20 | 2.41 | 1.23 |
| age | Age | 15.06 | 1.68 | 13.90 | 1.66 |
| female | 1 if the respondent is female | 0.53 | 0.50 | 0.55 | 0.50 |
| white | 1 if the respondent is white | 0.57 | 0.49 | 0.53 | 0.50 |
| born in the U.S. | 1 if born in the U.S. | 0.92 | 0.27 | 0.95 | 0.21 |
| years in school | Number of years in the current school | 2.57 | 1.44 | 3.09 | 1.74 |
| live with both parents | 1 if lives with both parents | 0.74 | 0.44 | 0.71 | 0.46 |
| *Parental Education* | | | | | |
| (less than HS) | 1 if parent's education is less than high school (HS) | 0.11 | 0.31 | 0.14 | 0.35 |
| HS grad | 1 if parent's education is HS or higher but no college degree | 0.45 | 0.50 | 0.47 | 0.50 |
| college grad | 1 if parent's education is college or higher | 0.32 | 0.47 | 0.23 | 0.42 |
| missing | 1 if parent's education information is missing | 0.12 | 0.32 | 0.15 | 0.36 |
| *Parent Job* | | | | | |
| (stay home) | 1 if parent is a homemaker, retired, or does not work | 0.09 | 0.29 | 0.12 | 0.33 |
| professional | 1 if parent's job is a doctor, lawyer, scientist, teacher, librarian, nurse, manager, executive, director | 0.28 | 0.45 | 0.20 | 0.40 |
| other jobs | 1 if parent's job is not "stay home" or "professional" | 0.54 | 0.50 | 0.56 | 0.50 |
| missing | 1 if parent's job information is missing | 0.09 | 0.29 | 0.11 | 0.32 |
| parental care | 1 if parent cares very much | 0.84 | 0.37 | 0.84 | 0.36 |
| neighborhood safety | 1 if feels safe in the neighbourhood | 0.74 | 0.44 | 0.73 | 0.44 |
| school safety | 1 if feels safe at school | 0.63 | 0.48 | 0.64 | 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 1: Estimation of Peer Effects in Study Effort

| | 2SLS | BC2SLS | GMM | BCGMM |
|---|---|---|---|---|
| local-aggregate peer effect | 0.0003 | 0.0002 | 0.0008 | 0.0007 |
| | (0.0015) | (0.0015) | (0.0015) | (0.0015) |
| local-average peer effect | 0.0311 | -0.0122 | 0.0839*** | 0.0779*** |
| | (0.0995) | (0.0570) | (0.0171) | (0.0167) |
| age | -0.0468*** | -0.0501*** | -0.0427*** | -0.0433*** |
| | (0.0102) | (0.0082) | (0.0071) | (0.0071) |
| female | 0.1444*** | 0.1434*** | 0.1465*** | 0.1466*** |
| | (0.0153) | (0.0152) | (0.0152) | (0.0152) |
| white | -0.0021 | -0.0018 | -0.0036 | -0.0036 |
| | (0.0188) | (0.0189) | (0.0188) | (0.0188) |
| born in the U.S. | -0.0374 | -0.0389 | -0.0368 | -0.0372 |
| | (0.0354) | (0.0353) | (0.0352) | (0.0352) |
| years in school | -0.0040 | -0.0041 | -0.0041 | -0.0041 |
| | (0.0055) | (0.0055) | (0.0055) | (0.0055) |
| live with both parents | 0.0019 | 0.0022 | 0.0013 | 0.0014 |
| | (0.0168) | (0.0168) | (0.0168) | (0.0168) |
| parent education: HS grad | 0.0483** | 0.0479** | 0.0486** | 0.0486** |
| | (0.0212) | (0.0212) | (0.0212) | (0.0212) |
| parent education: college grad | 0.0684*** | 0.0678*** | 0.0695*** | 0.0693*** |
| | (0.0251) | (0.0251) | (0.0250) | (0.0250) |
| parent education: missing | 0.1118*** | 0.1125*** | 0.1125*** | 0.1127*** |
| | (0.0275) | (0.0275) | (0.0274) | (0.0274) |
| parent job: professional | 0.0279 | 0.0265 | 0.0301 | 0.0297 |
| | (0.0265) | (0.0264) | (0.0264) | (0.0264) |
| parent job: other | -0.0010 | -0.0021 | 0.0005 | 0.0002 |
| | (0.0224) | (0.0223) | (0.0223) | (0.0223) |
| parent job: missing | 0.0835*** | 0.0822*** | 0.0847*** | 0.0843*** |
| | (0.0308) | (0.0307) | (0.0306) | (0.0306) |
| parental care | 0.1964*** | 0.1965*** | 0.1967*** | 0.1968*** |
| | (0.0208) | (0.0209) | (0.0208) | (0.0208) |
| neighborhood safety | 0.0773*** | 0.0774*** | 0.0784*** | 0.0784*** |
| | (0.0173) | (0.0174) | (0.0173) | (0.0173) |
| school safety | 0.0922*** | 0.0929*** | 0.0902*** | 0.0906*** |
| | (0.0165) | (0.0165) | (0.0165) | (0.0165) |
| contextual effects | Yes | Yes | Yes | Yes |
| network fixed effects | Yes | Yes | Yes | Yes |

Standard errors in parentheses. Statistical significance: ***p<0.01; **p<0.05; *p<0.1.
2SLS First Stage F test statistic: 9.360
2SLS OIR test p-value: 0.277
J test p-value for the null “local-aggregate model”: 0.00
J test p-value for the null “local-average model”: 0.66

Table 2: Estimation of Peer Effects in Sport Activities

| | 2SLS | BC2SLS | GMM | BCGMM |
|---|---|---|---|---|
| local-aggregate peer effect | 0.0057** | 0.0066*** | 0.0074*** | 0.0072*** |
| | (0.0026) | (0.0026) | (0.0025) | (0.0024) |
| local-average peer effect | 0.3095*** | 0.1038 | 0.1094*** | 0.1056*** |
| | (0.1237) | (0.0771) | (0.0182) | (0.0180) |
| age | -0.0216 | -0.0368*** | -0.0365*** | -0.0368*** |
| | (0.0158) | (0.0140) | (0.0129) | (0.0129) |
| female | -0.6707*** | -0.6828*** | -0.6824*** | -0.6827*** |
| | (0.0287) | (0.0280) | (0.0277) | (0.0277) |
| white | 0.0975*** | 0.1009*** | 0.1006*** | 0.1010*** |
| | (0.0346) | (0.0344) | (0.0343) | (0.0343) |
| born in the U.S. | 0.0950 | 0.1023 | 0.1011 | 0.1021 |
| | (0.0647) | (0.0642) | (0.0641) | (0.0641) |
| years in school | 0.0229** | 0.0195* | 0.0194* | 0.0194* |
| | (0.0103) | (0.0101) | (0.0100) | (0.0100) |
| live with both parents | -0.0036 | -0.0015 | -0.0015 | -0.0014 |
| | (0.0309) | (0.0306) | (0.0306) | (0.0306) |
| parent education: HS grad | -0.0181 | -0.0198 | -0.0194 | -0.0195 |
| | (0.0389) | (0.0387) | (0.0387) | (0.0387) |
| parent education: college grad | 0.0814* | 0.0754* | 0.0764* | 0.0762* |
| | (0.0461) | (0.0457) | (0.0456) | (0.0456) |
| parent education: missing | -0.0639 | -0.0647 | -0.0642 | -0.0643 |
| | (0.0503) | (0.0500) | (0.0500) | (0.0500) |
| parent job: professional | 0.0747 | 0.0792* | 0.0790 | 0.0795* |
| | (0.0485) | (0.0481) | (0.0481) | (0.0481) |
| parent job: other | 0.1060*** | 0.1096*** | 0.1101*** | 0.1103*** |
| | (0.0409) | (0.0406) | (0.0406) | (0.0406) |
| parent job: missing | 0.0363 | 0.0443 | 0.0438 | 0.0442 |
| | (0.0564) | (0.0559) | (0.0559) | (0.0559) |
| parental care | 0.0512 | 0.0485 | 0.0481 | 0.0480 |
| | (0.0383) | (0.0380) | (0.0380) | (0.0380) |
| neighborhood safety | 0.0933*** | 0.0996*** | 0.0998*** | 0.0999*** |
| | (0.0321) | (0.0317) | (0.0316) | (0.0316) |
| school safety | 0.0164 | 0.0141 | 0.0140 | 0.0141 |
| | (0.0303) | (0.0300) | (0.0300) | (0.0300) |
| contextual effects | Yes | Yes | Yes | Yes |
| network fixed effects | Yes | Yes | Yes | Yes |

Standard errors in parentheses. Statistical significance: ***p<0.01; **p<0.05; *p<0.1.
2SLS First Stage F test statistic: 6.881
2SLS OIR test p-value: 0.681
J test p-value for the null “local-aggregate model”: 0.00
J test p-value for the null “local-average model”: 0.00

Table 3: GMM Estimation of Local-Aggregate and Local-Average Models

| | Study Effort | | Sport Activities | |
|---|---|---|---|---|
| local-aggregate peer effect | 0.0025* | | 0.0133*** | |
| | (0.0014) | | (0.0022) | |
| local-average peer effect | | 0.0854*** | | 0.1320*** |
| | | (0.0166) | | (0.0166) |
| age | -0.0492*** | -0.0430*** | -0.0421*** | -0.0382*** |
| | (0.0070) | (0.0071) | (0.0128) | (0.0128) |
| female | 0.1457*** | 0.1456*** | -0.6893*** | -0.6818*** |
| | (0.0152) | (0.0152) | (0.0277) | (0.0277) |
| white | -0.0041 | -0.0024 | 0.1028*** | 0.1026*** |
| | (0.0188) | (0.0188) | (0.0344) | (0.0344) |
| born in the U.S. | -0.0411 | -0.0355 | 0.1081* | 0.1003 |
| | (0.0352) | (0.0352) | (0.0642) | (0.0642) |
| years in school | -0.0048 | -0.0037 | 0.0155 | 0.0218** |
| | (0.0055) | (0.0055) | (0.0100) | (0.0100) |
| live with both parents | 0.0023 | 0.0015 | 0.0005 | -0.0022 |
| | (0.0168) | (0.0168) | (0.0307) | (0.0307) |
| parent education: HS grad | 0.0471** | 0.0488** | -0.0213 | -0.0180 |
| | (0.0212) | (0.0212) | (0.0387) | (0.0387) |
| parent education: college grad | 0.0664*** | 0.0692*** | 0.0699 | 0.0796* |
| | (0.0250) | (0.0250) | (0.0457) | (0.0456) |
| parent education: missing | 0.1142*** | 0.1109*** | -0.0647 | -0.0640 |
| | (0.0274) | (0.0274) | (0.0501) | (0.0500) |
| parent job: professional | 0.0246 | 0.0297 | 0.0811* | 0.0832* |
| | (0.0264) | (0.0263) | (0.0481) | (0.0481) |
| parent job: other | -0.0036 | 0.0005 | 0.1126*** | 0.1126*** |
| | (0.0223) | (0.0222) | (0.0406) | (0.0406) |
| parent job: missing | 0.0800*** | 0.0852*** | 0.0483 | 0.0459 |
| | (0.0306) | (0.0306) | (0.0559) | (0.0559) |
| parental care | 0.1956*** | 0.1964*** | 0.0422 | 0.0515 |
| | (0.0208) | (0.0208) | (0.0380) | (0.0380) |
| neighborhood safety | 0.0777*** | 0.0773*** | 0.1029*** | 0.1006*** |
| | (0.0173) | (0.0173) | (0.0317) | (0.0316) |
| school safety | 0.0921*** | 0.0915*** | 0.0102 | 0.0177 |
| | (0.0165) | (0.0164) | (0.0301) | (0.0300) |
| contextual effects | Yes | Yes | Yes | Yes |
| network fixed effects | Yes | Yes | Yes | Yes |

Standard errors in parentheses. Statistical significance: ***p<0.01; **p<0.05; *p<0.1.

Table 4: Estimation of Peer Effects without Network Fixed Effects

| | Study Effort | | Sport Activities | |
|---|---|---|---|---|
| | 2SLS | GMM | 2SLS | GMM |
| local-aggregate peer effect | 0.0004 | 0.0009 | 0.0059*** | 0.0085*** |
| | (0.0014) | (0.0014) | (0.0025) | (0.0023) |
| local-average peer effect | 0.0833 | 0.1024*** | 0.3387*** | 0.1316*** |
| | (0.0866) | (0.0165) | (0.1065) | (0.0174) |
| age | -0.0519*** | -0.0494*** | -0.0167 | -0.0413*** |
| | (0.0107) | (0.0057) | (0.0160) | (0.0100) |
| female | 0.1451*** | 0.1466*** | -0.6681*** | -0.6821*** |
| | (0.0153) | (0.0152) | (0.0286) | (0.0277) |
| white | -0.0358** | -0.0358** | 0.1190*** | 0.1273*** |
| | (0.0181) | (0.0181) | (0.0332) | (0.0330) |
| born in the U.S. | -0.0251 | -0.0246 | 0.1164* | 0.1108* |
| | (0.0343) | (0.0338) | (0.0614) | (0.0616) |
| years in school | -0.0077 | -0.0077 | 0.0314*** | 0.0316*** |
| | (0.0052) | (0.0052) | (0.0095) | (0.0095) |
| live with both parents | -0.0038 | -0.0040 | 0.0055 | 0.0088 |
| | (0.0167) | (0.0168) | (0.0305) | (0.0306) |
| parent education: HS grad | 0.0375* | 0.0385* | -0.0274 | -0.0359 |
| | (0.0211) | (0.0212) | (0.0386) | (0.0385) |
| parent education: college grad | 0.0535** | 0.0544** | 0.0743 | 0.0587 |
| | (0.0248) | (0.0248) | (0.0457) | (0.0451) |
| parent education: missing | 0.1068*** | 0.1078*** | -0.0679 | -0.0698 |
| | (0.0273) | (0.0274) | (0.0498) | (0.0500) |
| parent job: professional | 0.0165 | 0.0176 | 0.0466 | 0.0417 |
| | (0.0265) | (0.0263) | (0.0479) | (0.0479) |
| parent job: other | -0.0111 | -0.0101 | 0.0770* | 0.0738* |
| | (0.0223) | (0.0222) | (0.0404) | (0.0404) |
| parent job: missing | 0.0825*** | 0.0831*** | 0.0238 | 0.0268 |
| | (0.0307) | (0.0308) | (0.0559) | (0.0560) |
| parental care | 0.2151*** | 0.2151*** | 0.0507 | 0.0464 |
| | (0.0207) | (0.0209) | (0.0379) | (0.0380) |
| neighborhood safety | 0.0687*** | 0.0695*** | 0.0805*** | 0.0833*** |
| | (0.0171) | (0.0172) | (0.0313) | (0.0314) |
| school safety | 0.0898*** | 0.0890*** | 0.0250 | 0.0256 |
| | (0.0161) | (0.0162) | (0.0294) | (0.0295) |
| contextual effects | Yes | Yes | Yes | Yes |
| network fixed effects | No | No | No | No |

Standard errors in parentheses. Statistical significance: ***p<0.01; **p<0.05; *p<0.1.

Table 5: Estimation of Peer Effects in Undirected Networks

| | Study Effort | | | |
|---|---|---|---|---|
| | 2SLS | BC2SLS | GMM | BCGMM |
| local-aggregate peer effect | -0.0002 | -0.0008 | -0.0002 | -0.0003 |
| | (0.0010) | (0.0009) | (0.0009) | (0.0009) |
| local-average peer effect | 0.7159*** | 0.1666* | 0.1138*** | 0.0802*** |
| | (0.2180) | (0.0979) | (0.0156) | (0.0156) |
| J test p-value for the null “local-aggregate model”: 0.00 | | | | |
| J test p-value for the null “local-average model”: 0.70 | | | | |
| | Sport Activities | | | |
| | 2SLS | BC2SLS | GMM | BCGMM |
| local-aggregate peer effect | 0.0063*** | 0.0078*** | 0.0081*** | 0.0085*** |
| | (0.0017) | (0.0015) | (0.0015) | (0.0014) |
| local-average peer effect | 0.4979** | 0.0747 | 0.0607*** | 0.0459*** |
| | (0.2265) | (0.1002) | (0.0166) | (0.0164) |
| J test p-value for the null “local-aggregate model”: 0.00 | | | | |
| J test p-value for the null “local-average model”: 0.00 | | | | |

Standard errors in parentheses. Statistical significance: ***p<0.01; **p<0.05; *p<0.1.
The control variables are the same as in Tables 1-2