DISCUSSION PAPER SERIES

No. 9366

PEER EFFECTS: SOCIAL MULTIPLIER OR SOCIAL NORMS?

Xiaodong Liu, Eleonora Patacchini and Yves Zenou

LABOUR ECONOMICS

Centre for Economic Policy Research

www.cepr.org

Available online at: **www.cepr.org/pubs/dps/DP9366.asp**

ISSN 0265-8003

PEER EFFECTS: SOCIAL MULTIPLIER OR SOCIAL NORMS?

Xiaodong Liu, University of Colorado, Boulder
Eleonora Patacchini, Syracuse University, EIEF and CEPR
Yves Zenou, Stockholm University, Research Institute of Industrial Economics (IFN), GAINS and CEPR

Discussion Paper No. 9366
February 2013

Centre for Economic Policy Research
77 Bastwick Street, London EC1V 3PZ, UK
Tel: (44 20) 7183 8801, Fax: (44 20) 7183 8820
Email: cepr@cepr.org, Website: www.cepr.org

CEPR Discussion Paper No. 9366

February 2013

ABSTRACT

Peer Effects: Social Multiplier or Social Norms?*

We develop an unified model embedding different behavioral mechanisms of social interactions and design a statistical model selection test to discriminate between them in empirical applications. This framework is applied to study peer effects in education and delinquent behavior for adolescents in the United States. We find that there are strong social multiplier effects in crime while, for education, social norms matter the most. This suggests that, for crime, individual-based policies are more appropriate while, for education, group-based policies are more effective.

JEL Classification: A14, D85 and Z13
Keywords: group-based policy, individual-based policy, social networks and spatial autoregressive model

Xiaodong Liu
Department of Economics
University of Colorado at Boulder
256 UCB
Boulder, Colorado 80309-0256
USA

Email: xiaodong.liu@colorado.edu

For further Discussion Papers by this author see:
www.cepr.org/pubs/new-dps/dplist.asp?authorid=173030

Eleonora Patacchini
Center for Policy Research
426 Eggers Hall
Syracuse University
Syracuse, NY 13244-1020
USA

Email: epatacch@maxwell.syr.edu

For further Discussion Papers by this author see:
www.cepr.org/pubs/new-dps/dplist.asp?authorid=158584

Yves Zenou
Department of Economics
Stockholm University
10691 Stockholm
SWEDEN

Email: yves.zenou@ne.su.se

For further Discussion Papers by this author see:
www.cepr.org/pubs/new-dps/dplist.asp?authorid=126027

*We thank Maria Marta Ferreyra, Benjamin Golub, Lung-fei Lee and Marc Sommer for very helpful 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.

Submitted 19 February 2013

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 consider two *social network models* aiming to capture the different ways peer effects operate.[1] In the *local-aggregate model*, 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. *Social multiplier* effects are clearly important in this model. In the *local-average model*, 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] We then develop a unified framework, the *hybrid network model*, which encompasses both local-aggregate and local-average effects.

We characterize the Nash equilibrium of each model and show under which condition an interior Nash equilibrium exists and is unique. Even though the two first models are fundamentally different in terms of behavioral foundation, it turns out that their implications in terms of Nash equilibrium are relatively close since only the adjacency matrix (which keeps track to whom each individual is connected) differs in equilibrium, one being the row-normalized version of the other. For the third model (hybrid network model), we are also able to characterize the equilibrium and show under

[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.

which condition there exists a unique Nash equilibrium. To the best of our knowledge, this is the first time such a model has been analyzed in the literature.

We then propose econometric counterparts of the theoretical models. 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 and not in the latter. Our theoretical analysis provides a microfoundation for these two models by showing under which condition and what kind of utility function is needed to obtain each of this model. In fact, we show that the equation studied by spatial econometricians for each model exactly corresponds to the best-reply function of each individual when they choose the effort in some activity (study for education, crime intensity for crime, etc.) that maximizes their utility function by taking the network as given. Depending on which type of utility considered, we obtain different results. In particular, if social norms matter when deciding how much effort to exert in some activity, then the best-reply function corresponds exactly to the local-average model while, if social multiplier effects are important, then it corresponds to the local-aggregate model.

In the second part of the paper, we give the conditions under which each model is identified. Most research in spatial econometrics and network econometrics has been performed for the local-average model, i.e. when the adjacency matrix is row-normalized. In the present paper, we study the case of a non-row-normalized adjacency matrix, which corresponds to our local-aggregate model. It turns out that the condition under which the local-aggregate model is identified is weaker than the local-average model. We also provide examples of networks where the local-aggregate model is identified while the local-average model is not. We also give the condition under which the hybrid network model can be identified.

In the third part, we extend Kelejian's (2008) J test for spatial econometric models to differentiate between the local-aggregate and the local-average effects in a social-interaction model with network fixed-effects. We also propose a hybrid model encompassing both local-aggregate and local-average effects and develop appropriate IV-based estimators. The traditional 2SLS estimator does not work well in our empirical analysis as the first-stage F test suggests the available IVs are weak. To fix the weak IV problem, we follow Lee (2007a) by generalizing the 2SLS estimator to a GMM estimator with additional quadratic moment conditions based on the correlation structure of the error term in the reduced-form equation. The GMM approach is easy to implement, asymptotically efficient, and allows us to test the hypothesis that the network formation (i.e. the adjacency matrix) is exogenous (conditional on covariates and network fixed effects) using an over-identifying restriction test.

Finally, in the last part of the paper, we test our framework using the U.S. National Longitudinal Survey of Adolescent Health (AddHealth), which contains unique detailed information on friendship relationships among teenagers. Empirical tests of models of social interactions are quite problematic because of the well-know issues that render the identification and measurement of peer effects quite difficult: (i) reflection, which is a particular case of simultaneity (Manski, 1993) and (ii) endogeneity, which may arise for both peer self-selection and unobserved network-specific effects. Our econometric strategy utilizes the structure of the network as well as network fixed effects and high quality individual information to clearly identify the peer effects from the contextual and correlated effects. This approach for the identification of peer effects, i.e. the use of network fixed effects in combination with high quality data on social contacts, has been used in 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). 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. This is extremely reasonable in our case study where the networks are extremely small (on average composed by roughly 6 people).

We find that, for juvenile delinquency, students are mostly influenced by the aggregate activity of their friends (local-aggregate model) while, for education, we show that both the social multiplier effect and social norm effect matter, even though the magnitude is higher for the latter effect. This indicates that students tend to *conform* to the social norm of their friends in terms of effort in education (local-average model). 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.

There are several papers that have formalized the local-aggregate model using a network approach (see, in particular, Ballester et al., 2006, 2010; Bramoullé and Kranton, 2007; Galeotti et al., 2009) and have tested this model for education (Calvó-Armengol et al., 2009) and crime (Calvó-Armengol et al., 2005; Patacchini and Zenou, 2008; Liu et al., 2012). There are fewer papers that have explicitly modeled the local-average model (Glaeser and Scheinkman, 2003) and have tested it for education (Lin, 2010; Boucher et al., 2012) or crime (Patacchini and Zenou, 2012). Ghiglino and Goyal (2010) develop a theoretical model where they compare the local aggregate and the local average model 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, 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 (hybrid model) are incorporated in a network model.

We believe that it is important to be able to (consider and) disentangle between different behavioral peer-effect mechanisms because they imply different policy implications. 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* should be implemented in the context of this model like, for example, a school-based or a place-based policy. 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 other words, in the local-aggregate model there is a *learning process* from peers that is not true in the local-average model. In that case, individual-based policy should be implemented.[4]

To sum-up, the contributions of this paper to the literature are as follows:

(i) We compare the equilibrium implications of the local-average and the local-aggregate model;

(ii) We propose a new theoretical model (the hybrid network model) where both social norms and social multiplier effects are taken into account and show under which condition there exists a unique Nash equilibrium;

(iii) We provide the identification condition for the hybrid network model;

(iv) We generalize the J test of Kelejian and Piras (2011) to a network model with network fixed effects, which allows us to design a statistical model selection test to detect which behavioral mechanism better represents the data;

(v) We empirically test the three models (local-aggregate, local-average and hybrid network model) using the U.S. National Longitudinal Survey of Adolescent Health (AddHealth) to study peer effects in school performance and juvenile delinquency.

The rest of paper is organized as follows. Section 2 introduces the theoretical framework for the

[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] See our discussion in Sections 2.6 and 6.

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. Finally, Section 6 concludes and discusses the different policy implications of each model.

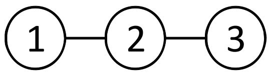
2 Theoretical Framework

2.1 The network

Suppose that a finite set of agents $N = \{1, \ldots, n\}$ is partitioned into $\overline{r}$ networks, where $N_r = \{1, \ldots, n_r\}$ denotes the set of agents in the rth network ($r = 1, ..., \overline{r}$). 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.[5] 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.[6] We denote by $g_r^{\max}$, the highest degree in network r, i.e. $g_r^{\max} = \max_i g_{i,r}$. If i and j are friends, then in general $N_{i,r} \neq N_{j,r}$ unless the network is complete. $N_{i,r} \cap N_{j,r} \neq \varnothing$ if i and j have common friends.

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 $0 \leq g_{ij,r}^* \leq 1$. Figure 1 gives an example of a network r 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 local-aggregate network model

In a network model, individuals simultaneously decide how much effort to exert in some activity to maximize their utility, given the effort levels chosen by their friends. First, we consider the case where an individual's utility depends on the *aggregate* effort level of her friends.

[5] 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.

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

We denote by $y_{i,r}$ the effort level of individual i in network r and by $Y_r = (y_{1,r}, ..., y_{n_r,r})'$ the population effort profile in network r. Given Y_r and the underlying network represented by G_r, individual i chooses an effort $y_{i,r} \geq 0$ to maximize her utility given by

$$u_{i,r}(y_{i,r}) \equiv u_{i,r}(y_{i,r}; Y_r, G_r) = (a_{i,r} + \eta_r + \epsilon_{i,r})\, y_{i,r} - \frac{1}{2} y_{i,r}^2 + \phi_1 \sum_{j=1}^{n_r} g_{ij,r} y_{i,r} y_{j,r}, \tag{1}$$

where $\phi_1 \geq 0$.

In the utility function (1), $(a_{i,r} + \eta_r + \epsilon_{i,r})\, y_{i,r}$ represents the benefit individual i received from effort $y_{i,r}$, where $a_{i,r} + \epsilon_{i,r}$ represents individual heterogeneity and η_r represents network heterogeneity of benefiting from a certain effort level. While $a_{i,r}$ is perfectly observable by all individuals in the network and the econometrician, η_r and $\epsilon_{i,r}$ are only observable by all individuals in the network but not by the econometrician. In (1), $\frac{1}{2} y_{i,r}^2$ captures the cost of effort. Note that the utility is concave in own decisions, and displays decreasing marginal returns in own effort levels.

The last term in (1) reflects the impact of friends' *aggregate* effort levels on i's utility. As individuals may have different locations in a network and their friends may choose different effort levels, the term $\sum_{j=1}^{n_r} g_{ij,r} y_{j,r}$ is heterogeneous in i. The coefficient ϕ_1 captures the *local-aggregate* endogenous peer effect. More precisely, bilateral influences for individual i, j $(i \neq j)$ are captured by the following cross derivatives,

$$\frac{\partial^2 u_{i,r}(y_{i,r})}{\partial y_{i,r} \partial y_{j,r}} = \phi_1 g_{ij,r}. \tag{2}$$

As we assume $\phi_1 > 0$, if i and j are friends, the cross derivative is positive and reflects strategic complementarity in efforts. That is, if j increases his effort, then the utility of i will be higher if i also increases her effort. Furthermore, the utility of i increases with the number of friends, given the effort profile Y_r.

In equilibrium, each agent maximizes her utility (1). From the first-order condition, we obtain the following best-reply function for individual i,

$$y_{i,r} = \phi_1 \sum_{j=1}^{n_r} g_{ij,r} y_{j,r} + a_{i,r} + \eta_r + \epsilon_{i,r}. \tag{3}$$

From (3), 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. Denote by $\pi_{i,r} = a_{i,r} + \eta_r + \epsilon_{i,r}$, $\Pi_r = (\pi_{1,r}, \cdots, \pi_{n_r,r})'$, and $g_r^{\max} = \max_i g_{i,r}$ the highest degree of network r. The following proposition characterizes the Nash equilibrium of the local-aggregate network model.[7]

Proposition 1 *If* $0 \leq \phi_1 g_r^{\max} < 1$*, then the network game with payoffs (1) has a unique interior Nash equilibrium in pure strategies given by (3). In matrix form, this can be written as:*

$$Y_r = (I_{n_r} - \phi_1 G_r)^{-1} \Pi_r. \tag{4}$$

2.3 The local-average network model

Let us now develop the *local-average* model where the *average effort level* of friends affects an individual's utility. Denote by $\overline{y}_{i,r} = \sum_{j \in N_{i,r}} g_{ij,r}^* y_{j,r}$ the average effort of individual i's friends. Given the population effort profile Y_r and the network structure G_r, individual i in network r selects an effort $y_{i,r} \geq 0$ and obtains a payoff given by

$$u_{i,r}(y_{i,r}) \equiv u_{i,r}(y_{i,r}; Y_r, G_r) = \left(a_{i,r}^* + \eta_r^* + \epsilon_{i,r}^*\right) y_{i,r} - \frac{1}{2} y_{i,r}^2 - \frac{d}{2}(y_{i,r} - \overline{y}_{i,r})^2 \tag{5}$$

with $d \geq 0$.

The first two terms of the utility function (5) have the same interpretation as in (1). Let us focus on the last term of this utility function since it is the only component that differs from (1). Indeed, the term $(y_{i,r} - \overline{y}_{i,r})^2$ reflects the influence of friends' behavior on one's own action. It is such that each individual wants to minimize the *social distance* between herself and her reference group. Its coefficient d describes the *taste for conformity.* Here, the individual loses utility $\frac{d}{2}(y_{i,r} - \overline{y}_{i,r})^2$ from failing to conform to others. 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).

Observe that beyond the individual heterogeneity $a_{i,r} + \epsilon_{i,r}$, the term $\overline{y}_{i,r}$ represents a second type of heterogeneity, referred to as the *peer heterogeneity*, which captures the differences between individuals due to network effects**.** Here it means that individuals have different types of friends and thus different reference groups $N_{i,r}$. As a result, the social norm each individual i faces is different.

In equilibrium, agents choose their effort level $y_{i,r} \geq 0$ simultaneously to maximize their utility. As $\overline{y}_{i,r} = \sum_{j \in N_{i,r}} g_{ij,r}^* y_{j,r}$ (remember that $g_{ij,r}^* = g_{ij,r}/g_{i,r}$), it follows from the first-order condition

[7] All proofs can be found in Appendix A.

that the best-reply function of individual i is given by

$$y_{i,r} = \phi_2 \sum_{j=1}^{n_r} g^*_{ij,r} y_{j,r} + a_{i,r} + \eta_r + \epsilon_{i,r}, \tag{6}$$

where $\phi_2 = d/(1+d)$, $a_{i,r} = (1-\phi_2)a^*_{i,r}$, $\eta_r = (1-\phi_2)\eta^*_r$, and $\epsilon_{i,r} = (1-\phi_2)\epsilon^*_{i,r}$. As the equilibrium effort level of individual i depends on the *average* effort of her friends, this model is referred to as the *local-aggregate* network model. The following proposition gives the Nash equilibrium of model.

Proposition 2 *If $0 \leq \phi_2 < 1$, then the network game with payoffs (5) has a unique interior Nash equilibrium in pure strategies given by (6). In matrix form, this can be written as:*

$$Y_r = (I_{n_r} - \phi_2 G^*_r)^{-1} \Pi_r. \tag{7}$$

2.4 The hybrid network model

Finally, we generalize the network model by integrating *local-aggregate* and *local-average* effects into the same model. Consider the following utility function

$$u_{i,r}(y_{i,r}) \equiv u_{i,r}(y_{i,r}; Y_r, G_r) = \left(a^*_{i,r} + \eta^*_r + \epsilon^*_{i,r}\right) y_{i,r} - \frac{1}{2} y^2_{i,r} + d_1 \sum_{j=1}^{n_r} g_{ij,r} y_{i,r} y_{j,r} - \frac{d_2}{2}(y_{i,r} - \overline{y}_{i,r})^2, \tag{8}$$

where $d_1 \geq 0$ and $d_2 \geq 0$. Thus, individual i' utility is positively affected by the sum of total efforts of her friends *and* negatively affected by the distance from the social norm of her reference group (i.e. the average effort of her friends). From the first-order condition of utility maximization, the best-reply function of individual i is then given by:

$$y_{i,r} = \phi_1 \sum_{j=1}^{n_r} g_{ij,r} y_{j,r} + \phi_2 \sum_{j=1}^{n_r} g^*_{ij,r} y_{j,r} + a_{i,r} + \eta_r + \epsilon_{i,r} \tag{9}$$

where $\phi_1 = d_1/(1+d_2)$, $\phi_2 = d_2/(1+d_2)$, $a_{i,r} = (1-\phi_2)a^*_{i,r}$, $\eta_r = (1-\phi_2)\eta^*_r$, and $\epsilon_{i,r} = (1-\phi_2)\epsilon^*_{i,r}$. As $d_1 > 0$ and $d_2 > 0$, we have $\phi_1 > 0$ and $0 < \phi_2 < 1$. This model is referred to as the *hybrid* network model and it includes the *local-aggregate* and *local-average* models as special cases. The Nash equilibrium of the hybrid network model is characterized by the following proposition.

Proposition 3 *If $\phi_1 \geq 0$, $\phi_2 \geq 0$ and $\phi_1 g_r^{\max} + \phi_2 < 1$, then the network game with payoffs (8) has a unique interior Nash equilibrium in pure strategies given by (9). In matrix form, this can be written as:*

$$Y_r = (I_{n_r} - \phi_1 G_r - \phi_2 G_r^*)^{-1} \Pi_r \tag{10}$$

2.5 Local aggregate versus local average: equilibrium comparison

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 utility. On the contrary, 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 utility. In the *local-aggregate* model, even if individuals are ex ante identical, different positions in the network would imply different equilibrium effort levels. However, in the *local-average* model, positions in the network would not matter and equilibrium effort levels would be the same if all individuals are ex ante identical.

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-aggregate model, if $0 \leq \phi_1 g_r^{\max} < 1$, the unique Nash equilibrium given by (4) now becomes

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

where l_{n_r} is an n_r-dimensional vector of ones. 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.[8] On the other hand, for the local-average model, if $0 \leq \phi_2 < 1$, the unique interior Nash equilibrium given by (7) now becomes[9]

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

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. This is one of the fundamental

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

[9] Note that $(I_{n_r} - \phi_2 G_r^*)^{-1} 1_{n_r} = (1-\phi_2)^{-1} 1_{n_r}$.

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

As a numerical example, consider the star-shaped network in Figure 1 with three ex ante identical individuals. Suppose $\pi_{1,r} = \pi_{2,r} = \pi_{3,r} = 1$. Then, for the *local-aggregate* model, if $0 \leq \phi_1 < 1/2$, individual 2, who is in the center of the network, will exert the equilibrium effort $y^*_{2,r} = \frac{1+2\phi_1}{1-2\phi_1^2}$ and individuals 1 and 3 will exert the equilibrium effort $y^*_{1,r} = y^*_{3,r} = \frac{1+\phi_1}{1-2\phi_1^2}$. Obviously, $y^*_{2,r} > y^*_{1,r} = y^*_{3,r}$. Consequently, individual 2 receives the highest equilibrium utility given by $u_{2,r}(y^*_{2,r}) \equiv \frac{1}{2}(\frac{1+2\phi_1}{1-2\phi_1^2})^2$, while individuals 1 and 3 receive the equilibrium utility $u_{1,r}(y^*_{1,r}) \equiv u_{3,r}(y^*_{3,r}) = \frac{1}{2}(\frac{1+\phi_1}{1-2\phi_1^2})^2$. On the other hand, for the local-average model, if $0 \leq \phi_2 < 1$, the equilibrium efforts are the same for all individuals such that $y^*_{1,r} = y^*_{2,r} = y^*_{3,r} = (1 - \phi_2)^{-1}$. Therefore, all three individuals receive the same equilibrium payoff.

Thus, these two models have fundamentally different equilibrium implications as illustrated in the above ex ante homogeneous 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 because they are ex ante identical.

2.6 Discussion

Let us discuss the differences between the local-aggregate and local-average model 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. Observe that, in the local-average model, the cost of deviating from the social norm depends on the size of the group the individual belongs to. This is important from a policy viewpoint. Indeed, if only one person deviates from a large group of friends, then the cost could be very high. Anson (1985) told the story of an African-American youth, Eddie Perry, who left his black poor neighborhood to enter into a very prestigious prep school but at the cost of considerable psychological pain, because he did not fit naturally into either his old world of the inner city or his new world of the prep school. In particular, when Eddie went back to his old neighborhood, his friends ridiculed him because he couldn't play basketball properly. As one of Eddie's mentors put it: "They ridiculed him for going away to school, they ridiculed him for turning white".

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* should be implemented in the context of this model like, for example, a school-based or a place-based policy. 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 other words, in the local-aggregate model there is a *learning process* from peers that is not true in the local-average model. In that case, individual-based policy should be implemented. We discuss the policy implications of these models in more details at the end of the paper in Section 6.

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" or "acting white" 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. This implies that, in a conformist group, changing people's behavior is much more difficult than in a "local-aggregate" group.

3 Identification of the Network Model

3.1 Econometric network models

The specification of the econometric models follows the equilibrium best-reply functions of the theoretical models exactly. This is important because it gives a clear microfoundation for each model. Let the ex ante heterogeneity $a_{i,r}$ of individual i in network r be

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

where $x_{i,r}$ is an m-dimensional vector of exogenous variables and β, γ are corresponding vectors of parameters. The econometric *local-aggregate* model corresponding to (3) is

$$y_{i,r} = \phi_1 \sum_{j=1}^{n_r} g_{ij,r} y_{j,r} + x'_{i,r}\beta + \sum_{j=1}^{n_r} g^*_{ij,r} x'_{j,r}\gamma + \eta_r + \epsilon_{i,r}, \tag{11}$$

the econometric *local-average* model corresponding to (6) is

$$y_{i,r} = \phi_2 \sum_{j=1}^{n_r} g^*_{ij,r} y_{j,r} + x'_{i,r}\beta + \sum_{j=1}^{n_r} g^*_{ij,r} x'_{j,r}\gamma + \eta_r + \epsilon_{i,r}, \tag{12}$$

and the econometric *hybrid* model corresponding to (9) is:

$$y_{i,r} = \phi_1 \sum_{j=1}^{n_r} g_{ij,r} y_{j,r} + \phi_2 \sum_{j=1}^{n_r} g^*_{ij,r} y_{j,r} + x'_{i,r}\beta + \sum_{j=1}^{n_r} g^*_{ij,r} x'_{j,r}\gamma + \eta_r + \epsilon_{i,r}, \tag{13}$$

for $i = 1, \cdots, n_r$ and $r = 1, \cdots, \bar{r}$. In the econometric models, $\epsilon_{i,r}$'s are i.i.d. innovations with zero mean and variance σ^2 for all i and 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, the econometric models can be written in matrix form as:

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

For a data set with $\bar{r}$ groups, let $Y = (Y_1', \cdots, Y_{\bar{r}}')'$, $X^* = (X_1^{*\prime}, \cdots, X_{\bar{r}}^{*\prime})'$, $\eta = (\eta_1, \cdots, \eta_{\bar{r}})'$, $\epsilon = (\epsilon_1', \cdots, \epsilon_{\bar{r}}')'$, $G = \text{diag}\{G_r\}_{r=1}^{\bar{r}}$, $G^* = \text{diag}\{G_r^*\}_{r=1}^{\bar{r}}$ and $L = \text{diag}\{l_{n_r}\}_{r=1}^{\bar{r}}$, where $\text{diag}\{A_r\}$ is a 'generalized' block diagonal matrix in which the diagonal blocks are $m_r \times n_r$ matrices A_r's. For the entire sample, the econometric models are, respectively,

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

$$Y = \phi_2 G^* Y + X\beta + G^* X\gamma + L\eta + \epsilon, \tag{15}$$

$$Y = \phi_1 GY + \phi_2 G^* Y + X\beta + G^* X\gamma + L\eta + \epsilon. \tag{16}$$

The econometric network models considered in this paper are quite general as they include *the endogenous effect*, where an individual's choice/outcome may depend on those of his/her friends, *the contextual effect*, where an individual's choice/outcome may depend on the exogenous characteristics of his/her friends, and *the network correlated effect* where individuals in the same network may behave similarly as they have similar unobserved individual characteristics or they face a similar institutional environment. Furthermore, we distinguish between *the aggregate endogenous effect*, captured by the coefficient ϕ_1, and *the average endogenous effect*, captured by the coefficient ϕ_2, as they originate from different economic models with totally different equilibrium implications.

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 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. If individuals i, j are friends and j, k are friends, it does not necessarily imply that i, k are also friends. Thus, the *intransitivity in social connections* provides an exclusion restriction to identify endogenous and contextual effects. Based on this important observation, Bramoullé et al. (2009) have derived identification conditions for the local-average model (only). In this section, we extend their results by giving the identification conditions of the local-aggregate model and the hybrid model and show that the aggregate endogenous effect is relatively easier to identify than the average endogenous effect.

The network correlated effect is captured by the vector of network fixed effects η. Network fixed effects can be interpreted as originating from a two-step model of link formation 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.

In the econometric models, we allow network fixed effects η to depend on G, G^* and X by treating η as a vector of unknown parameters (as in a fixed effect panel data model). When the number of groups $\bar{r}$ is large, we may have the incidental parameter problem. To avoid the incidental parameter problem, we transform (14)-(16) 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 fixed effect panel data model. As $JL = 0$, the transformed equations are

$$JY = \phi_1 JGY + JX\beta + JG^*X\gamma + J\epsilon, \tag{17}$$

$$JY = \phi_2 JG^*Y + JX\beta + JG^*X\gamma + J\epsilon, \tag{18}$$

$$JY = \phi_1 JGY + \phi_2 JG^*Y + JX\beta + JG^*X\gamma + J\epsilon. \tag{19}$$

The proposed estimators and identification conditions are based on the transformed models.

The transformed models can be estimated by generalizing the 2SLS and GMM methods in Liu and Lee (2010).[10] Let $Z_1 = [GY, X, G^*X]$, $Z_2 = [G^*Y, X, G^*X]$, and $Z_3 = [GY, G^*Y, X, G^*X]$ denote the matrices of regressors in the local-aggregate, local-average, and hybrid network models,

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

respectively. Let Q denote the IV matrix described below in Section 4.2. Then, the models are identified if the following condition is satisfied.

Identification Assumption $\lim_{n\to\infty} \frac{1}{n} Q' \mathrm{E}(JZ_k)$ is a finite matrix with full column rank for $k = 1, 2, 3$.

This identification assumption implies the rank condition that $\mathrm{E}(JZ_k)$ (for $k = 1, 2, 3$) has full column rank and that the column rank of Q at least as high as that of $\mathrm{E}(JZ_k)$, for large enough n. In the rest of this section, we provide sufficient conditions of the identification assumption for the models (17)-(19).

3.2 Identification of the local-average model

If $0 < \phi_2 < 1$, the reduced form equation of the local-average model (15) is

$$Y = (I - \phi_2 G^*)^{-1}(X\beta + G^* X\gamma + L\eta + \epsilon). \tag{20}$$

As $J(I - \phi_2 G^*)^{-1}L = 0$ and $(I - \phi_2 G^*)^{-1} = I + \phi_2 G^*(I - \phi_2 G^*)^{-1}$, it follows from (20) that

$$\mathrm{E}(JG^*Y) = JG^*X\beta + JG^*(I - \phi_2 G^*)^{-1} G^* X(\phi_2\beta + \gamma). \tag{21}$$

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 the matrix $\mathrm{E}(JZ_2) = [\mathrm{E}(JG^*Y), JX, JG^*X]$ does not have full column rank. This corresponds to the case where the endogenous effect and exogenous effect exactly cancel out. Lee et al. (2010) have shown, in this case, the reduced form equation of the local-average model becomes a simple regression model with (spatially) correlated disturbances. In the reduced form, there are neither endogenous nor contextual effects. Interactions go through unobservables (disturbances) instead of observables. A special case of $\phi_2\beta + \gamma = 0$ is $\beta = \gamma = 0$. In this case, $\mathrm{E}(JG^*Y) = 0$. The model cannot be identified as there is no relevant exogenous covariate in the model.

(ii) $\phi_2\beta + \gamma \neq 0$. In this case, the identification may still be hard to achieve when the network is dense. Lee (2007b) considered the case where individuals are equally influenced by all the others in the same network with the adjacency matrix $G_r^* = \frac{1}{n_r - 1}(l_{n_r} l_{n_r}' - I_{n_r})$. As $J_r G_r^* = \frac{1}{1 - n_r} J_r$ and

$J_r G_r^*(I_{n_r} - \phi_2 G_r^*)^{-1} G_r^* = [\sum_{j=0}^{\infty} \phi_2^j (\frac{1}{1-n_r})^{2+j}] J_r$, it follows from (21) that, for network r,

$$\begin{aligned} \mathrm{E}(J_r G_r^* Y_r) &= J_r G_r^* X_r \beta + J_r G_r^*(I_{n_r} - \phi_2 G_r^*)^{-1} G_r^* X_r(\phi_2 \beta + \gamma) \\ &= \frac{1}{1-n_r} J_r X_r \beta + [\sum_{j=0}^{\infty} \phi_2^j (\frac{1}{1-n_r})^{2+j}] J_r X_r (\phi_2 \beta + \gamma) \end{aligned}$$

and $J_r G_r^* X_r = \frac{1}{1-n_r} J_r X_r$. Therefore, $\mathrm{E}(JZ_2)$ does not have full column rank if all networks in the sample are of the same size such that $n_r = n/\bar{r}$. Lee (2007a) has shown that the model can be identified if there are variations in network sizes. However, the identification can be weak when all the networks are large. For a general local-average model, Bramoullé et al. (2009) and Lee et al. (2010) have derived some sufficient conditions for model identification. Suppose JX has full column rank. Bramoullé et al. (2009) have shown that if the matrices I, G^*, G^{*2}, G^{*3} are linearly independent, social effects are identified. For the perspective of a spatial autoregressive (SAR) model, Lee et al. (2010) have shown that the model can be identified if $J[X, G^*X, G^{*2}X]$ has full column rank.

As stated above, most research in spatial econometrics and network econometrics has been done for the local-average model, i.e., when the adjacency matrix is row-normalized and equal to G_r^*. In the next sections, we study the case of non row-normalized adjacency matrix G_r, which corresponds to our local-aggregate model and hybrid model.

3.3 Identification of the local-aggregate model

For network r, if $0 < \phi_1 g_r^{\max} < 1$, the reduced form equation of the local-aggregate model is

$$Y_r = (I_{n_r} - \phi_1 G_r)^{-1}(X_r \beta + G_r^* X_r \gamma + \eta_r l_{n_r} + \epsilon_r), \tag{22}$$

which implies

$$\mathrm{E}(J_r G_r Y_r) = J_r G_r (I_{n_r} - \phi_1 G_r)^{-1}(X_r \beta + G_r^* X_r \gamma) + \eta_r J_r G_r (I_{n_r} - \phi_1 G_r)^{-1} l_{n_r}. \tag{23}$$

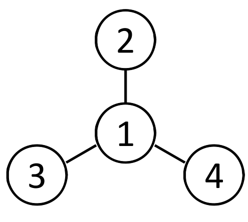
For network r, the term $G_r\left(I_{n_r} - \phi_1 G_r\right)^{-1} l_{n_r}$ is the Bonacich measure of centrality (Ballester et al. 2006; Bonacich, 1987). When row sums of G_r are not constant, the term $J_r G_r\left(I_{n_r} - \phi_1 G_r\right)^{-1} l_{n_r} \neq 0$ and therefore provides useful information for model identification. Even for the case that $\beta = \gamma = 0$, $\mathrm{E}(J_r Z_{1,r}) = [\mathrm{E}(J_r G_r Y_r), J_r X_r, J_r G_r^* X_r] = [\eta_r J_r G_r (I_{n_r} - \phi_1 G_r)^{-1} l_{n_r}, J_r X_r, J_r G_r^* X_r]$ may still have full column rank. As identification of the local-aggregate model requires $\mathrm{E}(JZ_1)$ 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.

Proposition 4

- *When G_r has non-constant row sums for some network r, $\mathrm{E}(JZ_1)$ of the local-aggregate network model (14) has full column rank if:* (i) $I_{n_r}, G_r, G_r^*, G_rG_r^*$ *are linearly independent and* $|\beta| + |\gamma| + |\eta_r| \neq 0$*; or* (ii) $G_rG_r^* = c_1I_{n_r} + c_2G_r + c_3G_r^*$ *and* Λ_1 *given by (36) has full rank.*
- *When G_r has constant row sums such that $g_{i,r} = g_r$ for all r, $\mathrm{E}(JZ_1)$ 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,* $GG^{*2} = c_1I + c_2G + c_3G^* + c_4GG^* + c_5G^{*2}$*, and* Λ_2 *given by (37) has full rank; or* (v) $g_r = g$ *for all r,* I, G^*, G^{*2}, G^{*3} *are linearly independent, and* $\phi_1\beta g + \gamma \neq 0$.

The identification conditions for the *local-aggregate* model given in Proposition 4 are weaker than the conditions for the *local-average* model in Bramoullé et al. (2009).

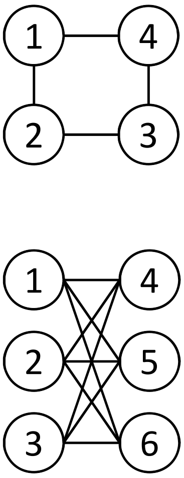
Interestingly, Figure 2 gives an example where identification is possible for the local-aggregate model but fails for the local-average model. Indeed, consider a dataset where each network is represented by the graph in Figure 2 (star-shaped network). The adjacency matrix G is a block-diagonal matrix 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 (15) is not identified. On the other hand, as G_r in Figure 2 has non-constant row sums and $I_{n_r}, G_r, G_r^*, G_rG_r^*$ are linearly independent, it follows from our Proposition 4(i) that the local-aggregate model (14) can be identified with this network.



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

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

Figure 3 provides another example where the local-average model cannot be identified while the local-aggregate model can. Indeed, consider a dataset with two types of networks. The first type of network is represented by the graph on the top of Figure 3 (regular network or circle). The second type of network is represented by the graph on the bottom of Figure 3 (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 (15) 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 4(iv).



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

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

Figure 3: An example where the local-aggregate model can be identified by Proposition 4(iv).

3.4 Identification of the hybrid model

Let us now consider the hybrid model and study identification. If $\phi_1 > 0$, $\phi_2 > 0$ and $\phi_1 g_r^{\max} + \phi_2 < 1$ for all r, the reduced form equation of the hybrid model (16) is

$$Y = (I - \phi_1 G - \phi_2 G^*)^{-1}(X\beta + G^* X\gamma + L\eta + \epsilon), \tag{24}$$

which implies

$$\begin{aligned}
\mathrm{E}(JGY) &= JG(I-\phi_1 G-\phi_2 G^*)^{-1}(X\beta+G^*X\gamma)+JG(I-\phi_1 G-\phi_2 G^*)^{-1}L\eta, &(25)\\
\mathrm{E}(JG^*Y) &= JG^*(I-\phi_1 G-\phi_2 G^*)^{-1}(X\beta+G^*X\gamma)+JG^*(I-\phi_1 G-\phi_2 G^*)^{-1}L\eta. &(26)
\end{aligned}$$

First, we consider the case that 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 hybrid model cannot be identified as $\mathrm{E}(JZ_3) = [\mathrm{E}(JGY), \mathrm{E}(JG^*Y), JX, JG^*X] = [g\mathrm{E}(JG^*Y), \mathrm{E}(JG^*Y), JX, JG^*X]$ does not have full column rank. If there are at least two networks in the data that have different degrees so that G and G^* are linearly independent, then the hybrid model can be identified through the following proposition.

Proposition 5 *Suppose G_r has constant row sums such that $g_{i,r} = g_r$ for all r. $\mathrm{E}(JZ_3)$ of the hybrid network model (16) 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 $\gamma + \phi_2\beta \neq 0$.*

This is a new result that allows for the identification of the hybrid model. Observe that if the row sums of G_r are not constant for some network r and $\eta_r \neq 0$, then $JG(I - \phi_1 G - \phi_2 G^*)^{-1}L$ can be used as an IV for the local-aggregate endogenous variable JGY. Furthermore, if $\phi_1 \neq 0$, then $JG^*(I - \phi_1 G - \phi_2 G^*)^{-1}L$ can be used as an IV for the local-average endogenous variable JG^*Y. Therefore, $\mathrm{E}(JZ_3)$ may have full column rank 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-aggregate model and the local-average model have different equilibrium implications (economic theory) and identification conditions (econometrics theory). However, their empirical specification (14) and (15) have similar functional forms, which allows us to design a statistical model selection test to detect which behavioral mechanism better represents the data.

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). The J test of Kelejian (2008) and Kelejian and Piras (2011) is implemented using the spatial 2SLS estimation procedure. This 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) The first-stage F test suggests that the IVs are weak in our data and thus the J test based on the 2SLS estimator would not be 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 (15) can be written more compactly as:

$$H_1 \quad : \quad Y = \phi_1 GY + X^*\delta_1 + L\eta_1 + \epsilon_1, \tag{27}$$

$$H_2 \quad : \quad Y = \phi_2 G^*Y + X^*\delta_2 + L\eta_2 + \epsilon_2, \tag{28}$$

where $X^* = (X, G^*X)$, and δ_1, δ_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,

$$Y = \alpha_1 Y_{H_2} + \phi_1 GY + X^* \delta_1 + L\eta_1 + \epsilon_1, \tag{29}$$

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 (27) against the alternative one (28) 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 (29) gives

$$\begin{aligned} Y &= \alpha_1(\phi_2 G^* Y + X^* \delta_2) + \phi_1 GY + X^* \delta_1 + L(\eta_1 + \alpha_1 \eta_2) + \epsilon_1 \\ &= Z_1^* \vartheta_1 + L(\eta_1 + \alpha_1 \eta_2) + \epsilon_1, \end{aligned} \tag{30}$$

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 (30) with the deviation from group mean projector J gives

$$JY = JZ_1^* \vartheta_1 + J\epsilon_1. \tag{31}$$

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 ϕ_2 and δ_2 be denoted by $\tilde{\phi}_2$ and $\tilde{\delta}_2$.

(2) Estimate the feasible counterpart of model (31)

$$JY = J\tilde{Z}_1^* \vartheta_1 + J\epsilon_1, \tag{32}$$

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 α_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,

$$H_2 : Y = \alpha_2 Y_{H_1} + \phi_2 G^* Y + X^* \delta_2 + L\eta_2 + \epsilon_2, \tag{33}$$

where Y_{H_1} is a predictor of Y under H_1 such that $Y_{H_1} = \phi_1 GY + X^*\delta_1 + L\eta_1$. Thus, the test of the null model (28) against the alternative (27) would be in terms of the hypotheses $H_0 : \alpha_2 = 0$ against $H_a : \alpha_2 \neq 0$. The within transformation of (33) with the projector J gives

$$JY = JZ_2^* \vartheta_2 + J\epsilon_2. \tag{34}$$

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 ϕ_1 and δ_1 be denoted by $\tilde{\phi}_1$ and $\tilde{\delta}_1$.

(2) Estimate the feasible counterpart of model (34)

$$JY = J\tilde{Z}_2^* \vartheta_2 + J\epsilon_2, \tag{35}$$

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 α_2 is significant, then that is evidence against model H_2.

4.2 The 2SLS and GMM estimators

For the estimation of the hybrid network model (19) and the augmented models (32) or (35) 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):

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

(b) "BC2SLS": a bias-corrected 2SLS estimator with IVs $Q_2 = J[X, G^*X, GX, G^{*2}X, GL]$. The additional IVs GL corresponds to the information on different positions of group members measured by Bonacich (1987) centrality. The additional IVs improves asymptotic efficiency of the estimator and helps achieve identification when the "conventional" IVs Q_1 are weak. Note that, the additional

IVs in Q_2, GL, 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 IVs Q_2 may have an asymptotic bias, which is known as the many-instrument bias.[11] 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 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.

5 Empirical Application

5.1 Data description

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,

[11] 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.

education, family background and friendship. This sample contains information on roughly 90,000 students.[12]

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. Pupils were asked to identify their best friends from a school roster (up to five males and five females).[13] Knowing exactly who nominates whom in a network, we exploit the directed nature of the nominations data.[14] 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.[15] 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.[16]

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 contains more than one network.

We exploit this unique data set to understand the impact of peer pressure on individual behavior for two different outcomes: (i) school performance and (ii) juvenile delinquency. We choose these two activities because they are very different from a behavioral perspective.

For each individual, we calculate an index of performance (or involvement) in each activity using answers to various related questions. More specifically, the *school performance* is measured using the respondent's scores received in the more recent grading period in several subjects, namely English or language arts, history or social science, mathematics and science. The scores are coded as 1=D or lower, 2=C, 3=B, 4=A. The final average score (*GPA index* or grade point average index) has

[12] 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 focus on wave I in-school data.

[13] 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.

[14] We also consider the undirected nature of the friendship relationships in Section 5.4.

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

[16] 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.

mean equal to 2.91 and standard deviation equal to 0.78.

The delinquency involvement is derived using questions describing participation in a series of activities that signal the propensity towards a delinquent behavior.[17] Each response is coded using an ordinal scale ranging from 0 (i.e. never participate) to 6 (i.e. nearly everyday). On the basis on these variables, a composite score is calculated for each respondent.[18] The final composite score (*delinquency index*) is normalized to be non-negative and has mean equal to 0.64 and standard deviation equal to 0.96. Precise definitions of the remaining variables (control variables) used in our empirical analysis can be found in Table C.1 in Appendix C. Excluding the individuals with missing or inadequate information, we obtain a final sample of 2,999 students distributed over 490 networks. The large reduction in sample size with respect to the original AddHealth sample is mainly due to missing values in variables and the network construction procedure. Indeed, roughly 20% of the students do not nominate any friends and another 20% cannot be correctly linked (for example because the identification code is missing or misreported). Also, we do not consider networks in the right tail of the network size distribution to avoid the possibility that in extremely large networks the strength of peer effects can be too different. Indeed, our theoretical model and hence our empirical strategy consider homogenous peer effects across networks. The use of network fixed effects, which is an important feature of our identification strategy prevents us from dealing with this issue. We focus our analysis on networks with a size of between 2 and 100 students for both outcomes.[19]

5.2 OIR test for the exogeneity of network structure

Our identification strategy is based on the assumption that any troubling source of heterogeneity, which is left unexplained by our 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. This is extremely reasonable in our case study where the networks are extremely small (the average network size is roughly 6 people, see Table C.1).

We will test the hypothesis that G is exogenous (conditional on covariates and network fixed effects) using an over-identifying restrictions (OIR) test, as described in Lee (2007a). The moment conditions used in the GMM estimation of a spatial autoregressive model (such as the ones used

[17] The in-school survey asks how often students have engaged in risky behavior during the past twelve months. Specifically, it contains information on how often students: smoked cigarettes, drank beer wine or liquor, got drunk, raced in bike or car, were in danger due to dare, lied to the parents, skipped school.

[18] We use a standard factor analysis, where the factor loadings of the different variables are used to derive the total score.

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

in our empirical investigation) are based on the assumption that G is exogenous. If the OIR test cannot reject the null hypothesis that the moment conditions (or restrictions) are correctly specified, then it provides evidence that G can be considered as exogenous. If the number of restrictions is small relative to the sample size, the OIR test statistic given by the GMM objective function (44) evaluated at the GMM estimator follows a chi-squared distribution with degrees of freedom equal to the number of over-identifying restrictions (Lee, 2007a, Proposition 2).

5.3 Estimation results

5.3.1 Statistical analysis

Let us now test our theoretical framework for different outcomes using the J test described above. Tables 1 and 2 report the estimation results for the hybrid network model using alternative estimators for education and crime, respectively. We consider the 2SLS and GMM estimators that are based on traditional moment conditions (i.e. 2SLS and GMM described in Section 4.2) and the bias-corrected version of 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 show the first stage F tests for weak IVs and the OIR tests in the lasts rows. In both tables, the values of the F-statistics reveal that our instruments in the linear moment conditions are quite weak.[20] As a result, we prefer the GMM with additional quadratic moment conditions (see Section 4.2 and Appendix B) to the 2SLS estimates. The other consequence of the weak IV problem is that we cannot trust the OIR test results based on the 2SLS. We thus consider the OIR test statistic (Lee, 2007a) for the GMM estimator.[21] We find that the p-values of the over-identifying restrictions test are large for all outcomes, which means that G_r is exogenous (i.e. we cannot reject the null hypothesis that the moment conditions based on an exogenous G_r are valid). This evidence provides confidence in *the exogeneity of network structure* (conditional on controls and network fixed effects).

[Insert Tables 1 and 2 here]

[20] The construction and statistical properties of the first stage F-test for weak instruments can be found in Stock et al. (2002). The partial F-test used in this analysis is further detailed in Stock and Yogo (2005). Stock et al. (2002) have suggested that "evidently the first-stage F statistic must be large, typically exceeding 10, for 2SLS inference to be reliable."

The matlab codes that implement the 2SLS and GMM estimators, J test, F test and OIR test remain available upon request.

[21] This test statistic uses the GMM estimator (GMM) based on a relatively small set of instruments (Q1). In fact, such a test could be biased if there are a large number of IVs, as in the IV matrix Q2. For the case with independent observations, Chao et al. (2010) have proposed an OIR test that is robust to many IVs. However, no robust OIR test with many IVs is available when observations are spatially correlated.

5.3.2 Interpretation of results in economics terms

Do peer effects matter? Which model is more adequate for each activity? These questions can be answered if we look at the first two rows of Tables 1 and 2 (tests of ϕ_1 and ϕ_2 of the hybrid network model) and at the last rows of these tables (tests of the augmented models where the null hypothesis is $\alpha_1 = 0$, i.e., the local average model does not matter for model (29), and $\alpha_2 = 0$, i.e., the local aggregate model does not matter for model (33).

Table 1 reveals that, for education (i.e. GPA index), both social norms (local average) and social multiplier (local aggregate) matter. However, even if both matter, the magnitude of the effects is higher for the local-average model compared to the local-aggregate one. Indeed, a one standard deviation increase in the average activity of individual i's reference group translates roughly into a 0.32 increase of a standard deviation of individual i's GPA score while it is only 0.13 for the sum of activity of friends. In a group of 4 friends, this means that if all my friends' GPA increase by 1 point, then my GPA would increase by 0.18 points because of my desire to conform to the social norm of the group and by about 0.06 because of the social multiplier effect. Social norms seem to be an important factor in peer effects in education.

On the contrary, in Table 2, we see clearly that, for juvenile delinquency, it is the *sum of the effort of the friends* (i.e. the local aggregate model) and not *their average effort* that matters for explaining own delinquent activity. In other words, there are *only multiplier peer effects* in crime activities. In terms of magnitude, a one standard deviation increase in the aggregate delinquent activity of individual i's reference group translates roughly into a 0.07 increase of a standard deviation of individual i's delinquency index.

Our results are interesting and new. First, they show that different forms of social interactions may drive peer effects in different outcomes. Second, they show that even for the same outcome 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. We report in Table 3 the results that are obtained when estimating separately the local-aggregate and the local-average model. Comparing Table 1 and 3, it appears clearly that for education (i.e. when more that a mechanism is at work) the local-average peer effect is overstated if the local-aggregate effect is ignored (and vice versa). In other words, if we estimate peer effects in education using a local-average model we would obtain a biased (and larger) estimate of the magnitude of those effects. Although there might be a variety of mechanisms at work, our framework is able to distinguish between the most important ones.

[Insert Table 3 here]

5.4 Robustness check

Our identification and estimation strategies depend on the correct specification of network links. In particular, our identification strategy hinges upon nonlinearities in group membership, 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 his/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 if both students have named each other, that is $g_{ij,r} = g_{ji,r} = 1$.

Tables 4 report the main results for undirected networks. The qualitative results remain unchanged.

[Insert Table 4 here]

6 Summary and policy implications

6.1 Summary

In this paper, we have studied three different social network models with different economic foundations. In the first one, the local-aggregate model, each individual is positively affected by the sum of effort of her friends (social multiplier) while, in the local-average model, it is the deviation from the average effort of her friends that has a negative impact on her utility (social norm). In the third model (the hybrid network model), local-aggregate as well as local-average effects are incorporated. We show that there exists a unique interior Nash equilibrium in each model.

We then propose different identification strategies for each model that we estimate using each individual's best-reply function. While the conditions under which models with row-normalized adjacency matrices are identified have been provided in the past, in this paper, we tackle the identification issue of non-row-normalized adjacency matrices. We give the exact conditions for identification and provide examples of networks that are identified under the local-aggregate model but not under the local-average one. We also propose an identification strategy for the hybrid network model. We then design a statistical model selection test to detect which behavioral mechanism better represents the data by generalizing the J test of Kelejian and Piras (2011) to a network model with network fixed effects.

In the last part of the paper, we test our framework using the AddHealth data, which provides detailed information on adolescent friendships in the United States. For juvenile delinquency, we find that students are mostly influenced by the sum of activities of their friends (local-aggregate model) while, for education, we show that both the aggregate school performance of friends and social norms matter, even though the magnitude of the effect is higher for the latter.

6.2 Policy implications

In this section, we would like to discuss the different policy implications of the local-aggregate, local-average and hybrid models. Indeed, we believe that it is important to be able to disentangle between different behavioral peer-effect models because they imply different policy implications. We base our discussion on the activities considered in our empirical application: crime and education. These are contexts where peer effects matter and where policy interventions are crucial.

6.2.1 Crime

It is well-documented that crime is, to some extent, a group phenomenon, and the source of crime is located in the intimate social networks of individuals (see e.g. Sutherland, 1947; Warr, 2002; Bayer et al., 2009). In the *local-aggregate model*, a key-player policy (Ballester et al., 2006; Liu et al., 2012), whose aim is to remove the criminal that reduces total crime in a network the most, would be the most effective policy since the effort of each criminal and thus the sum of one's friends crime efforts will be reduced. In other words, the removal of the key player can have large effects on crime because of the feedback effects or "social multipliers" at work (see, in particular, Kleiman, 2009; Glaeser et al., 1996; Verdier and Zenou, 2004). That is, as the fraction of individuals participating in a criminal behavior increases, the impact on others is multiplied through social networks. Thus, criminal behaviors can be magnified, and interventions can become more effective. On the contrary, a key-player policy would have nearly no effect in the *local-average model* since it will not affect the social norm of each group of friends in the network. To be effective, one would have to change the norm for each of the criminals, which is clearly a more difficult objective. In that case, one needs to target a group or gang of criminals to drastically reduce crime. This example of crime clear illustrates the fact that, for the local-aggregate model, individual-based policies are more appropriate while, for the local-average model, group-based policies are more effective. In our present analysis of juvenile deliquency, we have shown that the *local-aggregate model* was at work for the AddHealth data. This implies that a key-player policy would be the most effective policy to reduce crime for adolescents in the United States.

6.2.2 Education

Education is clearly an important topic and effective policies are difficult to implement. 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.[22] For example, in Section 2.6, 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

[22] 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, 2012).

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. 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.

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.

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] Anson, R.S. (1985), *Best Intentions: The Education and Killing of Edmund Perry*, New York: Random House.

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

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

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

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

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

[12] Bayer, P., Pintoff, R. and D. Pozen (2009), "Building criminal capital behind bars: Peer effects in juvenile corrections," *Quarterly Journal of Economics* 124, 105-147.

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

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

[15] 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.

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

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

[18] Boucher, V., Bramoullé, Y., Djebbari, H. and B. Fortin (2012), “Do peers affect student achievement? Evidence from Canada using group size variation,” *Journal of Applied Econometrics*, forthcoming.

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

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

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

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

[23] 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.

[24] 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.

[25] 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.

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

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

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

[29] De Martí, J. and Y. Zenou (2012), "Friendship formation, oppositional identity, and segregation," Unpublished manuscript, Stockholm University.

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

[31] 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.

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

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

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

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

[36] 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.

[37] Glaeser, E.L., Sacerdote, B. and J. Scheinkman (1996), "Crime and social interactions," *Quarterly Journal of Economics* 111, 508-548.

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

[39] 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.

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

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

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

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

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

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

[46] 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.

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

[48] 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.

[49] 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

[50] 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.

[51] 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.

[52] Kleiman, M.A. (2009), *When Brute Force Fails. How to Have Less Crime and Less Punishment*, Princeton: Princeton University Press.

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

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

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

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

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

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

[59] 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.

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

[61] 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/.

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

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

[64] 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.

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

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

[67] 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.

[68] 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.

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

[70] Verdier, T. and Y. Zenou (2004), "Racial beliefs, location and the causes of crime," *International Economic Review* 45, 727-756.

[71] Warr, M. (2002), *Companions in Crime: The Social Aspects of Criminal Conduct*, Cambridge: Cambridge University Press.

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

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

APPENDICES

A Proofs

Proof of Proposition 1. Apply Theorem 1, part b, in Calvó-Armengol et al. (2009) to our problem. In that case, the condition for existence and uniqueness is $\phi_1\omega_1(G_r) < 1$, where $\omega_1(G_r)$ is the spectral radius of G_r. As $\omega_1(G_r) < g_r^{\max}$ (Cvetković and Rowlinson, 1990, Theorem 1.1, page 4), a sufficient condition for $\phi_1\omega_1(G_r) < 1$ to be satisfied is $\phi_1 g_r^{\max} < 1$. ■

Proof of Proposition 2. Our utility function (5) can be written as:

$$\begin{aligned}
u_{i,r}(Y_r) &= \left(a_{i,r}^* + \eta_r^* + \epsilon_{i,r}^*\right) y_{i,r} - \frac{1}{2}y_{i,r}^2 - \frac{d}{2}(y_{i,r} - \overline{y}_{i,r})^2 \\
&= \left(a_{i,r}^* + \eta_r^* + \epsilon_{i,r}^*\right) y_{i,r} - \frac{(1+d)}{2}y_{i,r}^2 + dy_{i,r}\overline{y}_{i,r} - \frac{d}{2}\overline{y}_{i,r}^2 \\
&= \left(a_{i,r}^* + \eta_r^* + \epsilon_{i,r}^*\right) y_{i,r} - \frac{(1+d)}{2}y_{i,r}^2 + d\sum_{j=1}^{n_r} g_{ij,r}^* y_{i,r}y_{j,r} - \frac{d}{2}(\sum_{j=1}^{n_r} g_{ij,r}^* y_{j,r})^2.
\end{aligned}$$

We can now apply Theorem 1, part b, in Calvó-Armengol et al. (2009) to our problem.[23] The condition on eigenvalue (that guarantees that the Nash equilibrium is unique and interior) can now be written as: $1 + d > d\omega_1(G_r^*)$, or $1 > \phi_2\omega_1(G_r^*)$, as $\phi_2 = \frac{d}{1+d}$ and $d > 0$. Observe that G_r^* is a row-normalized matrix so that its largest eigenvalue is 1, i.e., $\omega_1(G_r^*) = 1$. As a result, if $\phi_2 < 1$ then the Nash equilibrium (6) is well-defined. ■

Proof of Proposition 3. We need to show that $I_{n_r} - A_r$ is invertible, where $A_r \equiv \phi_1 G_r + \phi_2 G_r^*$. A sufficient condition for the invertibility assumption can be derived as follows. Let $g_{i,r}$ be the ith row sum of G_r. Since G_r^* is the row-normalized G_r, we have $G_r = R_r G_r^*$, where R_r is a diagonal matrix with the ith diagonal element being $g_{i,r}$. The invertibility of $I_{n_r} - \phi_1 G_r - \phi_2 G_r^*$ requires that $||\phi_1 G_r + \phi_2 G_r^*|| < 1$ for some matrix norm $||\cdot||$ (see Horn and Johnson, 1900). Let $||\cdot||_\infty$ denote the row-sum matrix norm. As $||G_r^*||_\infty = 1$, we have that:

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

where $g_r^{\max} \equiv \max_i g_{i,r}$. Hence, a sufficient condition for the invertibility assumption would be that $\phi_1 g_r^{\max} + \phi_2 < 1$. ■

[23] Observe that the last term $-\frac{d}{2}(\sum_{j=1}^{n_r} g_{ij,r}^* y_{j,r})^2$ of the utility function does not matter since the derivative of this term with respect to $y_{i,r}$ is equal to zero.

Let

$$\Lambda_1 = \begin{bmatrix} \gamma c_1 & 1 & -\phi_1 c_1 \\ \beta + \gamma c_2 & -\phi_1 & -\phi_1 c_2 \\ \gamma c_3 & 0 & 1 - \phi_1 c_3 \\ \eta_r & 0 & 0 \end{bmatrix} \tag{36}$$

and

$$\Lambda_2 = \begin{bmatrix} -\gamma c_1 & 1 & \phi_1 c_1 \\ \beta - \gamma c_2 & -\phi_1 & \phi_1 c_2 \\ -\gamma c_3 & -1 & \phi_1 c_3 + 1 \\ \gamma - \beta - \gamma c_4 & \phi_1 & \phi_1 c_4 - \phi_1 \\ -\gamma c_5 & 0 & \phi_1 c_5 - 1 \end{bmatrix}. \tag{37}$$

Proof of Proposition 4. See Liu et al. (2012). ■

Proof of Proposition 5. 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

$$[\text{E}(JGY)d_1 + \text{E}(JG^*Y)d_2 + JXd_3 + JG^*Xd_4] = 0 \tag{38}$$

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 (25) and (26) that

$$\begin{aligned} \text{E}(JGY) &= J(I - \phi_1 G - \phi_2 G^*)^{-1}G(X\beta + G^*X\gamma), & (39) \\ \text{E}(JG^*Y) &= J(I - \phi_1 G - \phi_2 G^*)^{-1}G^*(X\beta + G^*X\gamma). & (40) \end{aligned}$$

Plugging (39) and (40) into (38) 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

$$Xd_3+GX(\beta d_1-\phi_1 d_3)+G^*X(\beta d_2-\phi_2 d_3+d_4)+GG^*X(\gamma d_1-\phi_1 d_4)+G^{*2}X(\gamma d_2-\phi_2 d_4) = \rho L, \tag{41}$$

for a constant scalar ρ. Premultiplying (41) by G^* gives

$$G^*Xd_3+GG^*X(\beta d_1-\phi_1 d_3)+G^{*2}X(\beta d_2-\phi_2 d_3+d_4)+GG^{*2}X(\gamma d_1-\phi_1 d_4)+G^{*3}X(\gamma d_2-\phi_2 d_4)=\rho L. \tag{42}$$

From (41) and (42), 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 hybrid 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 $\text{vec}_D(A) = (a_{11}, \cdots, a_{nn})'$, $A^s = A+A'$, $A^t = A-\text{tr}(A)J/\text{tr}(J)$, and A^- denote a generalized inverse of a square matrix A. For a parameter θ, let θ_0 denote the true parameter value in the data generating process. Let μ_3 and μ_4 denote, respectively, the third and fourth moments of the error term.

B.1 Estimation of the hybrid network model

B.1.1 2SLS estimation

Let $M_0 = (I - \phi_{10}G - \phi_{20}G^*)^{-1}$. From the reduced form equation (24), $\text{E}(Y) = M_0(X^*\delta_0 + L\eta)$.[24] For $Z = [GY, G^*Y, X^*]$, the ideal IV matrix for the explanatory variables JZ in (19) is given by

$$f = \text{E}(JZ) = J[G\text{E}(Y), G^*\text{E}(Y), X^*]. \tag{43}$$

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_∞ increases as the number of groups $\bar{r}$ increases. Furthermore, if $|\phi_{10}\max_i(\sum_j g_{ij})| + |\phi_{20}| < 1$,[25] 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_∞ can be approximated by a linear combination of $[I, G, G^*, G^2, GG^*, G^*G, G^{*2}, \cdots]$.

[24] 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.

[25] The model represents an equilibrium so $I - \phi_{10}G - \phi_{20}G^*$ is assumed to be invertible. In Proposition 3, 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 $|\phi_{10}|d^{\max} < 1$ (see Proposition 1) and a sufficient condition for the the invertibility of $I - \phi_{20}G^*$ for the local average model is $|\phi_{20}| < 1$ (see Proposition 2). 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 ξ_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 consider is $\hat{\theta}_{2sls} = (Z'P_KZ)^{-1}Z'P_KY$.

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 $b_{2sls} = \sigma^2(Z'P_KZ)^{-1}[\text{tr}(P_KGM_0), \text{tr}(P_KG^*M_0), 0_{1\times 2m}]' = O_p(K/n)$. 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 $\tilde{b}_{2sls} = \check{\sigma}^2(Z'P_KZ)^{-1}[\text{tr}(P_KGM), \text{tr}(P_KG^*M), 0_{1\times 2m}]'$ 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 ϵ at θ_0. As $\text{E}(\epsilon'U_1\epsilon) = \text{E}(\epsilon'U_2\epsilon) = 0$ 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

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

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

Schwarz inequality, the optimal GMM estimator is given by

$$\hat{\theta}_{gmm} = \arg\min g'(\theta)\Omega^{-1}g(\theta). \tag{44}$$

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

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

$\bar{D} = D - \frac{\mu_3}{\sigma^2}\omega' f$, and $\check{D} = D - \frac{\mu_3}{\sigma^2}\omega' P_K Z$. When $K^{3/2}/n \to 0$, the optimal GMM estimator[26] has the asymptotic distribution

$$\sqrt{n}(\hat{\theta}_{gmm} - \theta_0 - b_{gmm}) \xrightarrow{d} N(0, (\sigma^{-2}\bar{H} + \lim_{n\to\infty} \frac{1}{n}\bar{D}' B \bar{D})^{-1}), \tag{45}$$

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

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 θ_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 (45) 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.

[26] The weighting matrices for quadratic moments U_1, U_2 and the optimal weighting matrix for the objective function Ω^{-1} involves unknown parameters $\phi_1, \phi_2, \sigma_0^2, \mu_3$ and μ_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 (45).

B.2.1 2SLS estimation of the augmented model

First, we consider the 2SLS estimator of the augmented model (31). Let $M_{10} = (I - \alpha_{10}\phi_{20}G^* - \phi_{10}G)^{-1}$. The ideal IV matrix for JZ_1^* in (31) 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_∞ can be approximated by polynomials of I, G and G^*.

To achieve asymptotic efficiency, we consider an $n \times K$ feasible submatrix of Q_∞, 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[27] has the asymptotic distribution

$$\sqrt{n}(\hat{\vartheta}_{1,gmm} - \vartheta_{10} - b_{1,gmm}) \xrightarrow{d} N(0, (\sigma_1^{-2}\bar{H}_1 + \lim_{n\to\infty}\frac{1}{n}\bar{D}_1' B_1 \bar{D}_1)^{-1}), \tag{46}$$

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 (46).

[27] With consistent preliminary estimates of the unknown parameters in U_1, U_2, Ω, the feasible optimal GMM estimator can be shown to have the same asymptotic distribution given by (46).

Appendix C: Data appendix

Table C.1: Description of data

	Definition	Mean	SD
Dependent variables			
GPA	Average grade in math, science, English and history	2.91	0.78
delinquency index	In the text	0.64	0.96
Peer effects (in directed networks)			
average peer GPA	Average GPA of friends	2.31	1.35
average peer delinquency index	Average delinquency index of friends	0.52	0.77
aggregate peer GPA	Total GPA of friends	6.28	6.39
aggregate peer delinquency index	Total delinquency index of friends	1.21	1.76
Peer effects (in undirected networks)			
average peer GPA	Average GPA of friends	2.93	0.64
average peer delinquency index	Average delinquency index of friends	0.65	0.79
aggregate peer GPA	Total GPA of friends	9.78	8.99
aggregate peer delinquency index	Total delinquency index of friends	1.84	2.30
Control variables			
age	Age	14.21	1.80
female	1 if female; 0 otherwise	0.54	0.50
(white)	1 if white only; 0 otherwise	0.50	0.50
black	1 if black only; 0 otherwise	0.20	0.40
asian	1 if Asian only; 0 otherwise	0.04	0.19
other races	1 if race is not white, black or Asian; 0 otherwise	0.27	0.44
born in the U.S.	1 if born in the U.S.; 0 otherwise	0.90	0.31
years in school	Number of years in the current school	3.07	1.72
live with both parents	1 if live with both parents; 0 otherwise	0.71	0.45
parent education:	Dummy variable for parent's education. If both parents are in the household, the education of the father is considered.		
(less than HS)	1 if parent's education is less than high school (HS)	0.14	0.35
HS grad	1 if parent's education is HS or higher but no college degree	0.44	0.50
college grad	1 if parent's education is college or higher	0.27	0.44
missing	1 if parent's education information is missing	0.15	0.35
parent job:	Dummy variable for parent's job. If both parents are in the household, the job of the father is considered.		
(stay home)	1 if parent is a homemaker, retired, or does not work	0.10	0.30
professional	1 if parent's job is a doctor, lawyer, scientist, teacher, librarian, nurse, manager, executive, director	0.21	0.41
other jobs	1 if parent's job is not "stay home" or "professional"	0.57	0.49
missing	1 if parent's job information is missing	0.11	0.32
Network topology			
network size	Number of individuals in a network	6.12	14.44
row sum of directed adjacency matrix	Number of friends in the directed network	2.08	2.01
row sum of undirected adjacency matrix	Number of friends in the undirected network	3.26	2.83

Notes: The variable in the parentheses is the reference category. We consider networks with network size between 2 and 100 individuals.In our sample, there are 490 networks with 2999 total observations.

Table 1: 2SLS and GMM estimation of the hybrid model

School performance

	2SLS	BC2SLS	GMM	BCGMM
aggregate	0.0108***	0.0126***	0.0138***	0.0158***
	(0.0034)	(0.0031)	(0.0031)	(0.0031)
average	0.2188*	0.0921*	0.2193***	0.1843***
	(0.1183)	(0.0500)	(0.0340)	(0.0297)
age	-0.0424***	-0.0470***	-0.0389***	-0.0363***
	(0.0146)	(0.0139)	(0.0140)	(0.0139)
female	0.1980***	0.1913***	0.1992***	0.1964***
	(0.0307)	(0.0297)	(0.0301)	(0.0300)
black	-0.1351***	-0.1370***	-0.1424***	-0.1449***
	(0.0554)	(0.0546)	(0.0554)	(0.0552)
asian	0.0131	0.0375	0.0103	0.0347
	(0.1094)	(0.1061)	(0.1073)	(0.1068)
other races	-0.1276***	-0.1320***	-0.1351***	-0.1401***
	(0.0417)	(0.0410)	(0.0415)	(0.0414)
born in the U.S.	-0.0894	-0.0867	-0.0901	-0.0812
	(0.0598)	(0.0590)	(0.0598)	(0.0596)
years in school	0.0234**	0.0220**	0.0225**	0.0214**
	(0.0107)	(0.0105)	(0.0106)	(0.0106)
live with both parents	0.1602***	0.1680***	0.1598***	0.1660***
	(0.0348)	(0.0337)	(0.0341)	(0.0340)
parent education: HS grad	0.1431***	0.1471***	0.1465***	0.1493***
	(0.0439)	(0.0432)	(0.0438)	(0.0436)
parent education: college grad	0.2308***	0.2384***	0.2384***	0.2451***
	(0.0511)	(0.0500)	(0.0506)	(0.0504)
parent education: missing	0.1256**	0.1337***	0.1276**	0.1322***
	(0.0558)	(0.0546)	(0.0553)	(0.0551)
parent job: professional	0.1157**	0.1171**	0.1138**	0.1172**
	(0.0567)	(0.0559)	(0.0567)	(0.0565)
parent job: other	-0.0042	0.0005	-0.0054	-0.0000
	(0.0489)	(0.0481)	(0.0488)	(0.0486)
parent job: missing	0.0556	0.0530	0.0543	0.0566
	(0.0649)	(0.0640)	(0.0649)	(0.0647)
contextual effects	yes	yes	yes	yes
network fixed effects	yes	yes	yes	yes
1st stage F statistic	5.340			
OIR test p-value	0.655		0.137	
J test for the null hypothesis “local average model”: t statistic is 5.142				
J test for the null hypothesis “local aggregate model”: t statistic is 6.221				

Table 2: 2SLS and GMM estimation of the hybrid model

Delinquent activities

	2SLS	BC2SLS	GMM	BCGMM
aggregate	0.0227	0.0201	0.0325**	0.0391***
	(0.0221)	(0.0184)	(0.0151)	(0.0139)
average	0.5445***	0.1184*	0.0375	0.0283
	(0.2055)	(0.0620)	(0.0475)	(0.0406)
age	0.1049***	0.1170***	0.1174***	0.1205***
	(0.0204)	(0.0179)	(0.0177)	(0.0177)
female	-0.0982**	-0.0821**	-0.0804**	-0.0760**
	(0.0430)	(0.0386)	(0.0382)	(0.0382)
black	-0.1541**	-0.1511**	-0.1474**	-0.1570**
	(0.0780)	(0.0711)	(0.0704)	(0.0704)
asian	0.0789	0.0030	-0.0068	-0.0270
	(0.1551)	(0.1378)	(0.1364)	(0.1363)
other races	0.0865	0.0804	0.0812	0.0752
	(0.0586)	(0.0533)	(0.0528)	(0.0528)
born in the U.S.	0.0656	0.0561	0.0541	0.0618
	(0.0844)	(0.0768)	(0.0761)	(0.0761)
years in school	-0.0194	-0.0154	-0.0148	-0.0147
	(0.0151)	(0.0136)	(0.0135)	(0.0135)
live with both parents	-0.1139**	-0.1437***	-0.1462***	-0.1525***
	(0.0500)	(0.0439)	(0.0434)	(0.0433)
parent education: HS grad	-0.0665	-0.0950*	-0.0998*	-0.1034*
	(0.0631)	(0.0563)	(0.0557)	(0.0557)
parent education: college grad	-0.1335*	-0.1859***	-0.1940***	-0.2050***
	(0.0754)	(0.0652)	(0.0644)	(0.0644)
parent education: missing	-0.1752**	-0.2271***	-0.2349***	-0.2482***
	(0.0816)	(0.0712)	(0.0704)	(0.0703)
parent job: professional	0.0213	0.0185	0.0165	0.0152
	(0.0799)	(0.0728)	(0.0722)	(0.0721)
parent job: other	0.0296	0.0151	0.0122	0.0057
	(0.0690)	(0.0626)	(0.0620)	(0.0620)
parent job: missing	0.0699	0.0413	0.0409	0.0298
	(0.0924)	(0.0834)	(0.0826)	(0.0825)
contextual effects	yes	yes	yes	yes
network fixed effects	yes	yes	yes	yes
1st stage F statistic	2.252			
OIR test p-value	0.739		0.474	
J test for the null hypothesis “local average model”: t statistic is 2.804				
J test for the null hypothesis “local aggregate model”: t statistic is 0.698				

Table 3: GMM estimation of local-average and local-aggregate models

	School performance		*Delinquent activities*	
	GMM	BCGMM	GMM	BCGMM
aggregate		0.0222***		0.0457***
		(0.0029)		(0.0097)
average	0.3097***		0.6734***	
	(0.0007)		(0.0005)	
age	-0.0437***	-0.0422***	0.1020***	0.1189***
	(0.0141)	(0.0137)	(0.0201)	(0.0176)
female	0.2005***	0.1872***	-0.1031***	-0.0773**
	(0.0305)	(0.0295)	(0.0435)	(0.0381)
black	-0.1389***	-0.1429***	-0.1531*	-0.1490**
	(0.0562)	(0.0544)	(0.0802)	(0.0702)
asian	-0.0091	0.0756	0.0959	-0.0216
	(0.1086)	(0.1052)	(0.1552)	(0.1359)
other races	-0.1287***	-0.1440***	0.0909	0.0787
	(0.0421)	(0.0408)	(0.0602)	(0.0527)
born in the U.S.	-0.0931	-0.0760	0.0675	0.0562
	(0.0607)	(0.0588)	(0.0867)	(0.0759)
years in school	0.0269***	0.0180*	-0.0196	-0.0145
	(0.0108)	(0.0104)	(0.0154)	(0.0135)
live with both parents	0.1580***	0.1772***	-0.1066**	-0.1508***
	(0.0345)	(0.0335)	(0.0494)	(0.0432)
parent education: HS grad	0.1422***	0.1523***	-0.0613	-0.1038*
	(0.0444)	(0.0430)	(0.0634)	(0.0555)
parent education: college grad	0.2345***	0.2501***	-0.1206*	-0.2042***
	(0.0512)	(0.0497)	(0.0732)	(0.0641)
parent education: missing	0.1258**	0.1389***	-0.1637**	-0.2457***
	(0.0560)	(0.0543)	(0.0801)	(0.0701)
parent job: professional	0.1131**	0.1215**	0.0200	0.0137
	(0.0575)	(0.0557)	(0.0822)	(0.0719)
parent job: other jobs	-0.0070	0.0067	0.0313	0.0062
	(0.0494)	(0.0479)	(0.0706)	(0.0618)
parent job: missing	0.0538	0.0565	0.0746	0.0339
	(0.0658)	(0.0638)	(0.0941)	(0.0823)
contextual effects	yes	yes	yes	yes
network fixed effects	yes	yes	yes	yes

Table 4: Robustness check
2SLS and GMM estimation of the hybrid model
- *undirected networks* -

	School performance			
	2SLS	BC2SLS	GMM	BCGMM
aggregate	0.0073***	0.0068***	0.0101***	0.0097***
	(0.0023)	(0.0022)	(0.0020)	(0.0020)
average	0.5237***	0.6270***	0.1857***	0.2307***
	(0.1900)	(0.0970)	(0.0300)	(0.0278)
J test for the null hypothesis "local average model": t statistic is 4.939				
J test for the null hypothesis "local aggregate model": t statistic is 8.299				
	Delinquent activities			
	2SLS	BC2SLS	GMM	BCGMM
aggregate	0.0265*	0.0255**	0.0301***	0.0377***
	(0.0151)	(0.0126)	(0.0097)	(0.0092)
average	0.4858	0.1646*	0.0270	-0.0156
	(0.2961)	(0.0867)	(0.0402)	(0.0377)
J test for the null hypothesis "local average model": t statistic is 4.108				
J test for the null hypothesis "local aggregate model": t statistic is -0.414				

Note: The control sets is as in Tables 1-3