DISCUSSION PAPER SERIES

No. 9159

MULTIVARIATE CHOICE AND IDENTIFICATION OF SOCIAL INTERACTIONS

Ethan Cohen-Cole, Xiaodong Liu and Yves Zenou

LABOUR ECONOMICS

Centre for Economic Policy Research

www.cepr.org

Available online at: **www.cepr.org/pubs/dps/DP9159.asp**

ISSN 0265-8003

MULTIVARIATE CHOICE AND IDENTIFICATION OF SOCIAL INTERACTIONS

Ethan Cohen-Cole, University of Maryland
Xiaodong Liu, University of Colorado, Boulder
Yves Zenou, Stockholm University and CEPR

Discussion Paper No. 9159
September 2012

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. 9159

September 2012

ABSTRACT

Multivariate Choice and Identification of Social Interactions

In this paper, we investigate the impact of peers on own outcomes where all agents embedded in a network choose more than one activity. We develop a simple network model that illustrates these issues. We differentiate between the 'seemingly unrelated' simultaneous equations model where people are influenced only by others within the same activity, the 'triangular' simultaneous equations model, where there is some asymmetry in the peers' cross effects, and the 'square' simultaneous equations model, where all possible cross-choice effects are taken into account. We develop the conditions under which each model is identified, showing that the general 'square' simultaneous equations model with both simultaneity effect and cross-choice peer effect cannot be identified without any exclusion restrictions. We then study the impact of peer effects on education and screen activities and show that the estimated within- and cross-choice peer effects both have non-trivial impacts on adolescent behavior. We find, in particular, that, keeping peers' grades and screen activities fixed, watching more TV could be beneficial to a student's grade.

JEL Classification: C21, C3, I21 and Z13
Keywords: identification, peer effects and social networks

Ethan Cohen-Cole
Robert H Smith School of Business
University of Maryland, College Park
Van Munching Hall
College Park, MD
USA

Email: ecohencole@rhsmith.umd.edu

For further Discussion Papers by this author see:
www.cepr.org/pubs/new-dps/dplist.asp?authorid=163489

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

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

Submitted 25 September 2012

1 Introduction

Peer decisions and/or peer characteristics have been shown to be important in predicting different outcomes of individuals, ranging from education and crime to labor market (Ioannides and Loury, 2004; Sacerdote, 2011; Patacchini and Zenou, 2012). Most of this literature has, however, considered the effect of peers on one specific choice. For example, Calvó-Armengol et al. (2009) show that peers' grades affect own grades. Fletcher (2012) finds that a 10% increase in the proportion of classmates who drink increases the likelihood an individual drinks by five percentage points. Cutler and Glaeser (2010) find evidence for peer effects in smoking, etc. In reality, individuals make a multitude of choices, many of which of dependent on each other. As a result, peers can have multiple and sometimes opposite influences on their friends. For example, if a person's friends smoke, drink but perform well at school, how do these impact this person's choice across the three activities? This joint decision problem is what we study in the current paper. Our purpose is to help understand the impact of making more than one choice at a time; in particular, we discuss decision making with more than one choice in the context of peer influences and social networks. To the best of our knowledge, this is the first paper that analyzes this issue, at least from an econometric viewpoint.

To be more precise, our purpose in this paper is twofold. One, we show a set of methods for identification of peer influences in contexts of multiple decisions. Two, we show the empirical salience of these methods for education. In the empirics below, as well as many others that we explored, including more than a single decision changed the significance and magnitude of estimated peer effects. In some cases, the changes are qualitatively very important, including changing intuition about how peer effects in education operate in high school environments. Our conclusion from this finding is that empirical peer effects research should be carefully constructed to ensure that results reflect the full gamut of choices that individuals make.

Even with the large gains made to date in understanding how interactions influence individual decisions, a relatively large gap remains. To our knowledge, the existing literature on peer effects has entirely focused on a single decision by each individual. In analyzing the decision to study, we ignore the related and complex decisions to smoke, drink, play sports or commit crimes. While evaluating decisions in isolation makes the challenging empirics of social networks more tractable, it requires a host of unstated, and largely implausible, exclusion restrictions. Few people make important decisions about their life completely in isolation of other factors; however, this is precisely the assumption maintained by the literature.

The inclusion of more than one choice in a individual optimization problem introduces at least two complexities. The first is a standard simultaneity problem well known in the study of economics. Studying more means less time for sports and perhaps induces one to eat more, smoke less, etc. These generate a complementarity (or substitutability) of the decisions themselves, independent of social influences. Playing sports, studying, participating in other extracurriculars, and spending time with family all require large time commitments, which at some point reaches the maximum available time in a day. In this sense, these activities are substitutes. The presence of substitutability may mean that a student who plays three sports, and has friends who study much more than he/she does, will respond differently to the social influence of studying than another student that does not play sports. That is, he/she will either choose not to respond to the studying influence or will reduce his/her sports commitment. Estimation the studying decision in isolation of the sports decision means that we may incorrectly attribute his/her low responsiveness to the studying influence as the lack of social influence rather than the presence of other activities.

The second feature relates to the interdependence of social spillovers themselves and the link between the social effects of one action on the decision about the other action. For example, the decision to study is partially influenced by other's choices to study. The same fact is true about the decision to participate in sports. However, it also appears reasonable that the social influence of a student's peers on studying will also impact his/her decision to play sports. An encouragement to study more, whether tacit or direct, can lead to an influence on the decision to play sports. After having a successful study session, the student may go home and need to decide whether to play soccer for the spring. The social influence from the same group of friends can be quite distinct on separate decisions. This distinct influence is the channel from the social influence of friends playing soccer on studying and vice-versa. The other channel here is the determination in equilibrium of each of the two social influences. The relevance of social influence on studying can be quite different if students are also considering soccer at the same time, potentially reinforcing both effects, minimizing both or trading off between the two.

The current literature focuses on some single decision and the influence of peers on that decision (Ballester et al., 2006; Bramoullé and Kranton, 2007; Bramoullé et al., 2012). The spillover occurs as an optimization process in which agents enjoy utility by choosing similarly to their peers. Other actions, other groups, and other influences are assumed to be exogenous to this decision making process. Expanding to a set of two choices, in these models, individuals' optimization process would involve forming expectations in isolation over each activity. Deciding whether to study more assumes that the choices and the peer decisions relating to sports are already known. Similarly, in deciding whether to join the tennis team, the amount of studying and the choices of friends over studying are given. We relax this strong assumption. If others also choose activities jointly, the expectations that a student forms over peers' studying will depend on peers' sports decisions as well. Thus, in addition to the complementarity / substitutability associated with multiple decisions, there are transmission channels between choices that pass through the social process as well.

This paper will provide a structural model that helps to illustrate multivariate choice in a social interaction setting. As is normal in peer effects contexts, we begin from the view that individuals enjoy utility as a function of the actions of others. Next, we allow agents to choose more than one activity. The model will allow for these activities to have an arbitrarily degree of complementarity or substitutability. As well, the model is general enough to handle arbitrary combinations of choices; that is, we make no assumption of the orderings of choice bundles. Indeed, this generality is essential; combining sets of choices in a social interactions context into bundles dramatically restricts the set of possible actions available to individuals. It is easy to construct examples of preference reversals in the bundled goods setting that comply with standard choice axioms in the general setting here. We develop the general theoretical model in Section 2.1.

Section 3 begins our investigation of identification. Parallel to the economic drivers of the model are three identification issues. First, we have the same reflection problem (Manski 1993) that emerges from the coexistence of endogenous and contextual effects as in most social interactions studies.[1] Second, the social spillovers across choices adds a new level of difficulty. Finally, the simultaneity problem requires treatment as in many models. Here the social impacts make that problem more complex to solve. To better understand the layers of complexity added by these identification issues, we consider three models with increasing degrees of interdependence in decision making.

The first model we consider is the *'seemingly unrelated' simultaneous equations model*, where an individual's decision on an activity only depends on decisions of the peers on the same activity (the endogenous effect) and their characteristics (the contextual effect). The second model we consider is the *'triangular' system of simultaneous equations model*, where we introduce a one-way cross-choice

[1] One reason social networks matter in social interactions analysis is because they facilitate identification by breaking the reflection problem. This was originally recognized in Cohen-Cole (2006) and is systematically explored in Bramoullé et al. (2009).

peer effect besides the endogenous and contextual effects such that an individual's decision on an activity can be affected by the peers' decision on related activities. Finally, we consider the *'square' system of simultaneous equations model* with the simultaneity effect and two-way cross-choice peer effect so that an individual's decision on an activity can be affected by his/her own decision on related activities and the peers' decision on related activities. We allow the decision errors of an individual to be correlated across related activities in all three models. We show that the general 'square' simultaneous equations model with both simultaneity effect and cross-choice peer effect cannot be identified without any exclusion restrictions. To better understand this result, we consider three specifications of the 'square' model. In the first one, only cross-choice peer effects are considered. In the second one, only simultaneity effects are taken into account. The last specification is the *general 'square' model* where both simultaneity and cross-choice peer effects are considered. We show that the two first specifications can be identified *without* any exclusion restriction on the exogenous variables. In other words, it is the co-existence of those two effects that cause the non-identification of the general 'square' model. This is an interesting and somewhat surprising result since, usually, a simultaneous equations model with only simultaneity effect needs IV (that is an exclusion restrictions on the coefficients of exogenous variables) for identification. Here we show that, for a network model, this is not the case. All these identification issues are fully studied in Section 3.

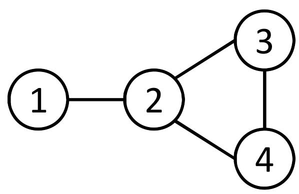
Section 4 proceeds with the empirical implementation of our models. In Section 4.1, we describe our data. We use the widely known Add-Health data for adolescents in the United States. We then test the different models for two activities: education (i.e. grades) and time spent watching TV or videos or playing video games and expose the empirical results. We find, in particular, that, keeping peers' grades and screen activities fixed, watching more TV could be beneficial to a student's grade. We conclude in Section 5.

In Appendix A, we list the matrices whose rank conditions are used for the identification of the 'square' model. In Appendix B, we present a more general econometric network model with m choices and give details of the generalized spatial 2SLS and 3SLS estimators. Appendix C collects all the proofs for the identification results in Section 3. In Appendix D, we discuss the identification and estimation of the network model with network-specific fixed effects.

2 Theoretical Model

2.1 A general model

There is finite set of agents $N = \{1, ..., n\}$ connected by a network $g \equiv g_N$. We keep track of social connections in a network g through its adjacency matrix $\boldsymbol{G} = [g^*_{ij}]$, where $g^*_{ij} = g_{ij} / \sum_{j=1}^n g_{ij}$ with $g_{ij} = 1$ if i and j are friends and $g_{ij} = 0$ otherwise.[2] We set $g_{ii} = 0$. Observe that $\boldsymbol{G}$ is a row-normalized matrix such that each row of $\boldsymbol{G}$ sums to one. An example is given in Figure 1 for a tree network with four agents.



$$\boldsymbol{G} = \begin{bmatrix} 0 & 1 & 0 & 0 \\ \frac{1}{3} & 0 & \frac{1}{3} & \frac{1}{3} \\ 0 & \frac{1}{2} & 0 & \frac{1}{2} \\ 0 & \frac{1}{2} & \frac{1}{2} & 0 \end{bmatrix}$$

Figure 1: an example network and the corresponding adjacency matrix.

As stated in the Introduction, agents choose effort levels for more than one activity. We denote by $\boldsymbol{y}_k = (y_{k1}, \cdots, y_{kn})'$ the vector of effort levels for the kth activity of agents in the network. Choices are made from elements of some set of possible choices (effort levels), Ω_k. This set is both individual- and network-specific. For every individual i, we must track the choices of other agents in the network.

We express the model with a linear quadratic utility function. Consider a structural model where individual i in the network chooses m actions to maximize his/her utility. For the ease of presentation, we focus on $m = 2$ without loss of generality. Let

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

where y_{1i} and y_{2i} are individual i's efforts in activities 1 and 2. As in the standard linear-quadratic utility model with one activity (Ballester et al., 2006), the first five terms of the utility function correspond to the impact of own characteristics on own effort levels while the last four terms express

[2] For ease of presentation, we assume that the network is undirected and no agent is isolated so that $\sum_{j=1}^n g_{ij} \neq 0$ for all i. The identification results of the paper hold for directed networks.

the impact of direct peers on own efforts. Apart from the fact that there are two activities and different level of heterogeneities, the main difference with the standard model is that there are new cross-effects, which are expressed by the parameters β_{12}^*, λ_{21}^* and λ_{12}^*.[3] The standard cross-effects between own and peer efforts for the same activity are

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

indicating strategic substitutability or complementarity depending on the signs of λ_{11}^* and λ_{22}^*. The new cross-effects are between own efforts for different activities, i.e.,

$$\frac{\partial^2 u(y_{1i}, y_{2i})}{\partial y_{1i} \partial y_{2i}} = \frac{\partial^2 u(y_{1i}, y_{2i})}{\partial y_{2i} \partial y_{1i}} = \beta_{12}^*, \tag{3}$$

and between own and peer efforts for different activities, i.e.,

$$\frac{\partial u(y_{1i}, y_{2i})}{\partial y_{1i} \partial y_{2j}} = g_{ij}^* \lambda_{21}^* \text{ and } \frac{\partial u(y_{1i}, y_{2i})}{\partial y_{2i} \partial y_{1j}} = g_{ij}^* \lambda_{12}^*, \tag{4}$$

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

The first order conditions of utility maximization are given by:

$$\begin{aligned} \frac{\partial u(y_{1i}, y_{2i})}{\partial y_{1i}} &= \alpha_{1i}^* - \beta_1^* y_{1i} + \beta_{12}^* y_{2i} + \lambda_{11}^* \textstyle\sum_{j=1}^n g_{ij}^* y_{1j} + \lambda_{21}^* \textstyle\sum_{j=1}^n g_{ij}^* y_{2j} = 0, && (5) \\ \frac{\partial u(y_{1i}, y_{2i})}{\partial y_{2i}} &= \alpha_{2i}^* - \beta_2^* y_{2i} + \beta_{12}^* y_{1i} + \lambda_{22}^* \textstyle\sum_{j=1}^n g_{ij}^* y_{2j} + \lambda_{12}^* \textstyle\sum_{j=1}^n g_{ij}^* y_{1j} = 0. && (6) \end{aligned}$$

From (5) and (6), we have the equilibrium *best-reply functions* as:

$$\begin{aligned} y_{1i} &= \alpha_{1i} + \beta_{12} y_{2i} + \lambda_{11} \textstyle\sum_{j=1}^n g_{ij}^* y_{1j} + \lambda_{21} \textstyle\sum_{j=1}^n g_{ij}^* y_{2j}, && (7) \\ y_{2i} &= \alpha_{2i} + \beta_{21} y_{1i} + \lambda_{22} \textstyle\sum_{j=1}^n g_{ij}^* y_{2j} + \lambda_{12} \textstyle\sum_{j=1}^n g_{ij}^* y_{1j}, && (8) \end{aligned}$$

where $\alpha_{1i} = \alpha_{1i}^*/\beta_1^*$, $\beta_{12} = \beta_{12}^*/\beta_1^*$, $\lambda_{11} = \lambda_{11}^*/\beta_1^*$, $\lambda_{21} = \lambda_{21}^*/\beta_1^*$, $\alpha_{2i} = \alpha_{2i}^*/\beta_2^*$, $\beta_{21} = \beta_{12}^*/\beta_2^*$,

[3] Because, in the empirical part, we focus on the identification of peer effects for the *row-normalized* adjacency matrix $\boldsymbol{G}$, our utility function also has this feature. In terms of economic interpretation, this means that it is the average behavior of peers (local-average model) that matters for own effort and not the sum of their efforts (local-aggregate model). In terms of utility function, an interpretation of the local-average model is that individuals *conform* to the average behavior of their peers and that deviating from this *social norm* could be costly (Liu, Patacchini and Zenou, 2012; Patacchini and Zenou, 2012).

$\lambda_{22} = \lambda_{22}^*/\beta_2^*$, and $\lambda_{12} = \lambda_{12}^*/\beta_2^*$. In Appendix B, we give a condition (Assumption 1) for the uniqueness of the equilibrium. This condition is analogous to the assumption that $(\boldsymbol{I} - \phi \boldsymbol{G})$ is invertible for the single activity network model (Ballester et al., 2006).

This theoretical model helps us understand agents' behavior when making more than one choice. First, this model helps analyze how a standard network model of social interactions can be generalized. Second, it provides economic fundamentals for the empirical work below. This will be very helpful when we will interpret our empirical results.

2.2 More specific models

Before moving to econometric analysis, we present a bit more on the behavioral foundation of the specific models we study.

2.2.1 The 'seemingly unrelated' simultaneous equations model

Consider a model where $\beta_{12}^* = \lambda_{12}^* = \lambda_{21}^* = 0$. In that case, the utility function can be written as:

$$u(y_{1i}, y_{2i}) = \alpha_{1i}^* y_{1i} + \alpha_{2i}^* y_{2i} - \frac{1}{2}\beta_1^* y_{1i}^2 - \frac{1}{2}\beta_2^* y_{2i}^2 + \lambda_{11}^* \textstyle\sum_{j=1}^n g_{ij}^* y_{1i} y_{1j} + \lambda_{22}^* \sum_{j=1}^n g_{ij}^* y_{2i} y_{2j}. \tag{9}$$

This is a model where only the peer-effects (2) within the same activity matter while the cross-effects (3) and (4) between the two activities are not taken into account.[4] In that case, equilibrium best-reply functions are equal to:

$$y_{1i} = \alpha_{1i} + \lambda_{11} \textstyle\sum_{j=1}^n g_{ij}^* y_{1j}, \tag{10}$$

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

This is very close to the model of Ballester et al. (2006) where people are influenced only by others within the same activity. Suppose y_{1i} is the time spent studying and y_{2i} is the time spent watching TV for individual i. Then, keeping α_{1i} (say, family size, parental education, etc.) fixed, individual i will spend more time studying (y_{1i}) if his/her friends also spend more time studying (if $\lambda_{11} > 0$). Similarly, keeping α_{2i} fixed, individual i will spend more time watching TV (y_{2i}) if his/her friends also watch more TV (if $\lambda_{22} > 0$). In this formulation, it is assumed that individual i choice on studying time is independent of her choice on TV time or the time that individual $i's$

[4] This is a model that has been considered from a theoretical viewpoint by Belhaj and Deroïan (2012) for the specific case when $y_{1i} + y_{2i} = 1$, $\forall i$, $\lambda_{11}^* > 0$ and $\lambda_{22}^* > 0$.

friends watch, and vice versa. This form of the model is conformable to the assumption that TV time choices are exogenous to the studying decision, but this is clearly a very strong assumption.

2.2.2 The 'triangular' simultaneous equations model

Let us now consider a utility function for which $\beta^*_{12} = \lambda^*_{12} = 0$. We have:

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

In this formulation, we still have peer effects between own efforts for different activities, i.e. (2), but we have an asymmetry for the cross effects between own and peer efforts for different activities. In other words, it is assumed that $\frac{\partial^2 u(y_{1i}, y_{2i})}{\partial y_{1i} \partial y_{2i}} = \frac{\partial^2 u(y_{1i}, y_{2i})}{\partial y_{2i} \partial y_{1i}} = \frac{\partial u(y_{1i}, y_{2i})}{\partial y_{2i} \partial y_{1j}} = 0$. In this case, the best-reply functions are equal to:

$$
\begin{aligned}
y_{1i} &= \alpha_{1i} + \lambda_{11} \textstyle\sum_{j=1}^n g^*_{ij} y_{1j} + \lambda_{21} \sum_{j=1}^n g^*_{ij} y_{2j}, && (12) \\
y_{2i} &= \alpha_{2i} + \lambda_{22} \textstyle\sum_{j=1}^n g^*_{ij} y_{2j}. && (13)
\end{aligned}
$$

If we keep the same interpretation as above, keeping α_{1i} fixed, the more time his/her friends spend studying (if $\lambda_{11} > 0$) and the less time his/her friends spend watching TV (if $\lambda_{21} < 0$), then the more time individual i will spend studying (y_{1i}). On the other hand, keeping α_{2i} fixed, the more his/her friends watch TV (if $\lambda_{22} > 0$) the more individual i will watch TV (y_{2i}). In this formulation, it is assumed that the time spent doing homework is influenced by how much time one's friends are watching TV while the reverse is not true, i.e. the time spent watching TV is not directly affected by how much time one's friends spend doing homework.

2.2.3 The 'square' simultaneous equations model

For the 'square' system of simultaneous equations, we consider three specifications.

First, we consider a utility function for which $\beta^*_{12} = 0$, so that (1) becomes

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

In this case, the best-reply functions are

$$y_{1i} = \alpha_{1i} + \lambda_{11} \sum_{j=1}^{n} g_{ij}^{*} y_{1j} + \lambda_{21} \sum_{j=1}^{n} g_{ij}^{*} y_{2j}, \tag{14}$$

$$y_{2i} = \alpha_{2i} + \lambda_{22} \sum_{j=1}^{n} g_{ij}^{*} y_{2j} + \lambda_{12} \sum_{j=1}^{n} g_{ij}^{*} y_{1j}, \tag{15}$$

For this specification, we assume how much time individual i will spend studying or watching TV would depend on how much time his/her friends spend studying and watching TV. However, this specification does not allow individual i's choice in one activity to be directly affected by his own choice in a related activity.

For the second specification, we let $\lambda_{12}^{*} = \lambda_{21}^{*} = 0$ in the utility function (1) so that

$$\begin{aligned} u(y_{1i}, y_{2i}) &= \alpha_{1i}^{*} y_{1i} + \alpha_{2i}^{*} y_{2i} - \frac{1}{2}\beta_{1}^{*} y_{1i}^{2} - \frac{1}{2}\beta_{2}^{*} y_{2i}^{2} + \beta_{12}^{*} y_{1i} y_{2i} \\ &\quad + \lambda_{11}^{*} \sum_{j=1}^{n} g_{ij}^{*} y_{1i} y_{1j} + \lambda_{22}^{*} \sum_{j=1}^{n} g_{ij}^{*} y_{2i} y_{2j}. \end{aligned}$$

In this case, the best-reply functions are

$$y_{1i} = \alpha_{1i} + \beta_{12} y_{2i} + \lambda_{11} \sum_{j=1}^{n} g_{ij}^{*} y_{1j}, \tag{16}$$

$$y_{2i} = \alpha_{2i} + \beta_{21} y_{1i} + \lambda_{22} \sum_{j=1}^{n} g_{ij}^{*} y_{2j}. \tag{17}$$

Suppose, as above, that y_{1i} is the time spent studying and y_{2i} is the time spent watching TV for individual i. Then, for this specification, how much time individual i will spend studying would depend on how much time he/she spends watching TV and how much time his/her friends spend studying. Similarly, how much time individual i will spend watching TV would depend on how much time he/she spends studying and how much time his/her friends spend watching TV. However, this specification assumes that how much time individual i will spend studying is not affected by how much time his/her friends spend watching TV, and vice versa.

Finally, we consider the most general model where we don't impose any restriction on the parameters. In that case, we end up with the utility function (1), where the equilibrium best-reply functions are given by (7) and (8). This is clearly the most interesting case since all activities influence each other. In other words, my time spent studying is affected by how much time I spend watching TV and by how much time my friends spend studying and watching TV. Similarly, my time spent watching TV is also affected by how much time I spend studying and by how much time my friends spend studying and watching TV.

2.3 Model Identification Challenges

As with most models in the peer effects literature, a host of identification issues emerge. Here, we are concerned both with the reflection problem for each choice that is endemic in peer effect research as well as with the complexities that emerge in estimating two or more equations simultaneously.

Our specification of the econometric model follow closely from the equilibrium best response functions (7) and (8) for the general theoretical model. Indeed, denote $\boldsymbol{x}_i = (x_{k_1 i}, \cdots, y_{k_x i})'$ a $1 \times k_x$ vector of observable individual-specific characteristics, $\sum_{j=1}^{n} g_{ij}^{*} \boldsymbol{x}_j$, a $1 \times k_x$ vector of observable network-specific characteristics (i.e. contextual effects), ϵ_{ki}, a random individual-specific characteristics associated with i, and a_k, a network-specific fixed effect. Let $\boldsymbol{c}_k$ and $\boldsymbol{\gamma}_k$ be two $k_x \times 1$ vectors of parameters. Then, α_{ki} in (7) and (8) can be written as

$$\alpha_{ki} = \boldsymbol{x}_i \boldsymbol{c}_k + \textstyle\sum_{j=1}^{n} g_{ij}^{*} \boldsymbol{x}_j \boldsymbol{\gamma}_k + a_k + \epsilon_{ki},$$

for $k = 1, 2$. The econometric simultaneous equations model corresponding to (7) and (8) is then given by:

$$\begin{aligned} y_{1i} &= \beta_{12} y_{2i} + \lambda_{11} \textstyle\sum_{j=1}^{n} g_{ij}^{*} y_{1j} + \lambda_{21} \sum_{j=1}^{n} g_{ij}^{*} y_{2j} + \boldsymbol{x}_i \boldsymbol{c}_1 + \sum_{j=1}^{n} g_{ij}^{*} \boldsymbol{x}_j \boldsymbol{\gamma}_1 + a_1 + \epsilon_{1i}, \quad (18) \\ y_{2i} &= \beta_{21} y_{1i} + \lambda_{22} \textstyle\sum_{j=1}^{n} g_{ij}^{*} y_{2j} + \lambda_{12} \sum_{j=1}^{n} g_{ij}^{*} y_{1j} + \boldsymbol{x}_i \boldsymbol{c}_2 + \sum_{j=1}^{n} g_{ij}^{*} \boldsymbol{x}_j \boldsymbol{\gamma}_2 + a_2 + \epsilon_{2i}. \quad (19) \end{aligned}$$

Our interest is to identify and estimate the various effects in the model, which are:[5]

- **Endogenous effect and contextual effect**

The endogenous effect, where an individual's choice may depend on his/her peers' choices on the same activity, is captured by the coefficients λ_{11} and λ_{22}. The contextual effect, wherein an individual's choice may depend on his/her peers' exogenous characteristics, is captured by $\boldsymbol{\gamma}_1$ and $\boldsymbol{\gamma}_2$. The reflection problem (Manski, 1993) is well known and emerges from the coexistence of those two effects. In Manski's 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

[5] For very complete overviews of identification problems in peer effects and how to solve them, see Durlauf and Ioannides (2010), Blume et al. (2011) and Ioannides (2012),

the same group introduces a perfect collinearity between the endogenous effect and the contextual effect. Hence, those two effects cannot be separately identified in the linear-in-means model.

Among the many methods to resolve this problem, one can use the complexity of social networks to identify the peer effect. In a social network, the assumption is that individuals are no longer impacted evenly by the full population in the sample; instead, they are influenced by their friends or connections. Bramoullé et al. (2009) have studied this phenomenon and they have shown that one can identify the endogenous and contextual effects if intransitivities exist in a network so that $\boldsymbol{I}, \boldsymbol{G}, \boldsymbol{G}^2$ are linearly independent. Intuitively, if individuals i and j are friends and j and k are friends, it does not necessarily imply that i and k are also friends. Because of these intransitivities, the characteristics of an individual's indirect friends are not collinear with his/her own characteristics and the characteristics of his/her direct friends. Therefore, the characteristics of an individual's indirect friends can be used as instruments to identify the endogenous effect from the contextual effect.

The 'seemingly unrelated' simultaneous equations model only considers the endogenous effect and the contextual effect for each choice. Although we allow the choices for related activities to be correlated through unobservables (i.e. the error terms of simultaneous equations), the identification condition will be shown in Section 3 to be exactly the same as that in Bramoullé et al. (2009). While this does not yet address the core issue we study, it relaxes the assumption in baseline models in the literature that choices such as studying and smoking are uncorrelated. This has no impact on the Bramoullé et al. (2009) identification conditions as they simply use the mean of the reduced-form equation, which is unaffected by the cross-equation correlation in the error terms.

- **Cross-choice peer effect**

Of course, choices are not just correlated, but also linked in important ways. A central component of our paper is that peers' choices may impact an individual's decisions on related activities (e.g. more smoking by friends may lead to less or more studying by an individual). We model this through the cross-choice peer effect. The coefficients λ_{21} and λ_{12} precisely describe how the choices of peers over other activities influences an individual's decisions about a given activity.

We illustrate the additional layer of complication in identification that emerges with these terms. We do so by considering a 'triangular' model, with a one-way cross-choice peer effect besides endogenous and contextual effects. This intermediate step to the full system, but the limited 'triangular' system is useful in understanding the mechanisms at work.

We will show below that to identify the cross-choice peer effect in the 'triangular' model, we need to further exploit the set of exclusion restrictions from the intransitivities that exist in a natural network of friendships. In Section 3.3, we formally show that the 'triangular' model can be identified when $\boldsymbol{I}, \boldsymbol{G}, \boldsymbol{G}^2, \boldsymbol{G}^3$ are linearly independent.

This can be understood in two ways. One, in a simple sense, we need to identify one additional structural parameter. To do so, we need an additional exclusion restriction of some type. We draw then again on the fact that the social network has a larger number of intransitivities. Every intransitivity is in fact an available exclusion; most social networks will have a very large number of them. Using the $\boldsymbol{G}^2$ matrix is a simple and intuitive one as it a characterization of the second degree links in the network. $\boldsymbol{G}^3$ shows the third degree links.

Two, intuitively, we can consider a classroom example where we need to understand how the smoking choices of students in a classroom influence smoking decisions, how studying decisions influence studying and how smoking decisions of peers influences studying. The first two are the case well known in the literature and can be identified using the intransitivities in the $\boldsymbol{G}$ matrix, as represented in the $\boldsymbol{G}^2$ matrix. However, notice that if we again use the $\boldsymbol{G}^2$ matrix to identify the influence of smoking decisions on studying, we would be unable to distinguish between the influence of smoking on smoking from the influence of smoking on studying. In precisely the same way that the reflection problem prevents identification in a linear model, we cannot separate the influence of smoking on the two decisions without an additional exclusion restriction. We use the $\boldsymbol{G}^3$ matrix as this exclusion restriction.

Furthermore, we show in Section 3.4 that a more complicated 'square' model with endogenous, contextual and two-way cross-choice peer effects can also be identified through the exclusion restrictions due to the intransitivity of friendship.

- **Simultaneity effect**

The standard economic simultaneity effect can be seen in the β_{12} and β_{21} coefficients of our general simultaneous equations model. Thus, increasing y_{2i} has an incremental impact on the choice of y_{1i}, and vice versa. In the absence of social influence, the simultaneity problem is a well known problem for the identification of a simultaneous equations model. The usual remedy for this identification problem is to impose exclusion restrictions on the coefficients of exogenous variables

$\boldsymbol{x}_i$.

For the 'square' simultaneous equations model, we show in Section 3.4 that for the specification with only simultaneity effect (besides endogenous and contextual effects), the model can be identified through the intransitivities in the network structure. However, for the general 'square' model with both simultaneity and cross-choice peer effects, the intransitivities in the network structure would not be enough to identify the various social interaction effects. Hence, to achieve identification, we need to impose some exclusion restrictions on $\boldsymbol{x}_i$.

- **Network correlated effect**

The structure of the general simultaneous equations model is flexible enough to allow us to incorporate two types of correlated effects.

First, the individuals in the same network may behave similarly as they have similar unobserved individual characteristics and they face similar institutional environment. We call this type of correlated effect *the network correlated effect.* The network correlated effect is captured in our general simultaneous equations model by introducing a network-specific fixed effect a_k.

Network fixed effects can be interpreted as originating from a two-step link formation model, where individuals self-select into different networks in a first step with selection bias due to specific network characteristics and, then, in a second step, link formation takes place within networks based on observable individual characteristics only. Therefore, network fixed effects serve as a (partial) remedy for the bias that originates from the possible sorting of individuals into networks. To estimate the simultaneous equations model with network fixed effects, we use a within group estimator that eliminates the network fixed effects by subtracting the network average from the individual-level variables. The details of the estimator and identification results are presented in Appendix D.

- **Cross-choice correlated effect**

Second, the decisions of the same individual on related activities may be correlated. We call this type of correlated effect *the cross-choice correlated effect.* The cross-choice correlated effect is introduced by allowing the error terms ϵ_{ki}'s to be correlated across equations. As our identification results are based on the mean of reduce-form equations, they will not be affected by the across-equation correlation in the error term. However, for estimation efficiency, it is important to consider the correlation structure of error terms. Following Kelejian and Prucha (2004), we adopt a generalized spatial 3SLS (GS3SLS) estimator for the estimation of our model. The details of the estimator are given in Appendix B.3.

3 Identification of Econometric Model

3.1 General setup

Consider a data set containing n agents, partitioned into $\bar{r}$ networks such that each network has n_r agents ($r = 1, \cdots, \bar{r}$) and $\sum_{r=1}^{\bar{r}} n_r = n$.[6] Links between agents are captured by an $n_r \times n_r$ zero-diagonal row-normalized adjacency matrix $\boldsymbol{G}_r = [g^*_{ij,r}]$ defined as in the previous section.

The equilibrium best response functions (18) and (19) can be rewritten in matrix form as[7]

$$\boldsymbol{y}_{1,r} = \lambda_{11}\boldsymbol{G}_r\boldsymbol{y}_{1,r} + \beta_{12}\boldsymbol{y}_{2,r} + \lambda_{21}\boldsymbol{G}_r\boldsymbol{y}_{2,r} + \boldsymbol{X}_r\boldsymbol{c}_1 + \boldsymbol{G}_r\boldsymbol{X}_r\boldsymbol{\gamma}_1 + \boldsymbol{\epsilon}_{1,r} \tag{20}$$

$$\boldsymbol{y}_{2,r} = \lambda_{22}\boldsymbol{G}_r\boldsymbol{y}_{2,r} + \beta_{21}\boldsymbol{y}_{1,r} + \lambda_{12}\boldsymbol{G}_r\boldsymbol{y}_{1,r} + \boldsymbol{X}_r\boldsymbol{c}_2 + \boldsymbol{G}_r\boldsymbol{X}_r\boldsymbol{\gamma}_2 + \boldsymbol{\epsilon}_{2,r}. \tag{21}$$

In this model, for $k = 1, 2$, $\boldsymbol{y}_{k,r} = (y_{k,1}, \cdots, y_{k,n_r})'$ is an $n_r \times 1$ vector of cross sectional observations on the kth decision. $\boldsymbol{X}_r = (\boldsymbol{x}'_1, \cdots, \boldsymbol{x}'_{n_r})'$ is an $n_r \times k_x$ matrix of exogenous variables. $\boldsymbol{\epsilon}_{k,r} = (\epsilon_{k,1}, \cdots, \epsilon_{k,n_r})'$ is an $n_r \times 1$ vector of disturbances. We assume $(\boldsymbol{\epsilon}_{1,r}, \boldsymbol{\epsilon}_{2,r}) = \boldsymbol{V}_r\boldsymbol{\Sigma}^{1/2}$, where $\boldsymbol{V}_r = [v_{ki,r}]$ is an $n_r \times 2$ matrix of I.I.D. innovations with zero mean and unit variance and

$$\boldsymbol{\Sigma} = \begin{bmatrix} \sigma_1^2 & \sigma_{12} \\ \sigma_{12} & \sigma_2^2 \end{bmatrix}$$

is a positive semi-definite symmetric matrix. Thus, we allow the error terms of decisions made by the same agent to be correlated, which is captured by σ_{12}.

For all $\bar{r}$ networks in the sample, we have, for $k = 1, 2$, $\boldsymbol{y}_k = (\boldsymbol{y}'_{k,1}, \cdots, \boldsymbol{y}'_{k,\bar{r}})'$, $\boldsymbol{\epsilon}_k = (\boldsymbol{\epsilon}'_{k,1}, \cdots, \boldsymbol{\epsilon}'_{k,\bar{r}})'$, $\boldsymbol{X} = (\boldsymbol{X}'_1, \cdots, \boldsymbol{X}'_{\bar{r}})'$, and $\boldsymbol{G} = \mathrm{diag}\{\boldsymbol{G}_r\}_{r=1}^{\bar{r}}$. Then,

$$\boldsymbol{y}_1 = \lambda_{11}\boldsymbol{G}\boldsymbol{y}_1 + \beta_{12}\boldsymbol{y}_2 + \lambda_{21}\boldsymbol{G}\boldsymbol{y}_2 + \boldsymbol{X}\boldsymbol{c}_1 + \boldsymbol{G}\boldsymbol{X}\boldsymbol{\gamma}_1 + \boldsymbol{\epsilon}_1 \tag{22}$$

$$\boldsymbol{y}_2 = \lambda_{22}\boldsymbol{G}\boldsymbol{y}_2 + \beta_{21}\boldsymbol{y}_1 + \lambda_{12}\boldsymbol{G}\boldsymbol{y}_1 + \boldsymbol{X}\boldsymbol{c}_2 + \boldsymbol{G}\boldsymbol{X}\boldsymbol{\gamma}_2 + \boldsymbol{\epsilon}_2. \tag{23}$$

As $(\boldsymbol{\epsilon}_1, \boldsymbol{\epsilon}_2) = \boldsymbol{V}\boldsymbol{\Sigma}^{1/2}$, where $\boldsymbol{V}$ is an $n \times 2$ matrix of I.I.D. innovations with zero mean and

[6] For simplicity, we assume no agent is isolated.

[7] For ease of presentation, we focus on the model without the network fixed effect a_k. The identification results for the model with network fixed effects are collected in Appendix D.

unit variance, $\mathrm{E}(\boldsymbol{\epsilon}\boldsymbol{\epsilon}') = \boldsymbol{\Sigma} \otimes \boldsymbol{I}_n$ for $\boldsymbol{\epsilon} = (\boldsymbol{\epsilon}_1', \boldsymbol{\epsilon}_2')'$. Let $\boldsymbol{Z}_1 = [\boldsymbol{G}\boldsymbol{y}_1, \boldsymbol{y}_2, \boldsymbol{G}\boldsymbol{y}_2, \boldsymbol{X}, \boldsymbol{G}\boldsymbol{X}]$, $\boldsymbol{Z}_2 = [\boldsymbol{G}\boldsymbol{y}_2, \boldsymbol{y}_1, \boldsymbol{G}\boldsymbol{y}_1, \boldsymbol{X}, \boldsymbol{G}\boldsymbol{X}]$, and $\boldsymbol{Q}$ denote the IV matrix based on $\boldsymbol{G}, \boldsymbol{X}$ and their functions.[8] Then, the model is identified if the following condition is satisfied.

Identification Assumption $\lim_{n\to\infty} \frac{1}{n}\boldsymbol{Q}'\mathrm{E}(\boldsymbol{Z}_k)$ is a finite matrix which has full column rank for $k = 1, 2$.

This identification assumption implies the rank condition that $\mathrm{E}(\boldsymbol{Z}_k)$ has full column rank and that $\boldsymbol{Q}$ has a rank at least as high as $\mathrm{E}(\boldsymbol{Z}_k)$, for large enough n. In the rest of this section, we provide some sufficient conditions for the identification assumption to hold.

This simultaneous equation model not only incorporates the endogenous effect (through λ_{11} and λ_{22}) and the contextual effect (through $\boldsymbol{\gamma}_1$ and $\boldsymbol{\gamma}_2$) as in a standard single equation network model, but also includes the simultaneity effect (through β_{12} and β_{21}), wherein an individual's decision depends on his/her decisions on related activities, the cross-choice peer effect (through λ_{21} and λ_{12}), wherein an individual's decision depends on his/her peers' decisions on related activities, and the cross-choice correlated effect through the correlation in $\boldsymbol{\epsilon}_1$ and $\boldsymbol{\epsilon}_2$. The general model structure will allow us to illustrate how the various effects operate on a empirical basis, and to show the identification requirements for each effect.

In the following subsections, we articulate methods to identify three different versions of the general simultaneous equations model that amount to increasing degrees of interdependence in decision making. As we discussed in the previous Section, in terms of the parameters of the utility function (1), the models are, namely, (i) the 'seemingly unrelated' regression (SUR) case where $\beta^*_{12} = \lambda^*_{12} = \lambda^*_{21} = 0$, (ii) the 'triangular' case where $\beta^*_{12} = \lambda^*_{12} = 0$; and (iii) the "square" case where we do not impose any restrictions on the parameters in (7) and (8). Without loss of generality, henceforth we assume $k_x = 1$ so that there is a unique exogenous variable in the model as in Bramoullé et al. (2009).

3.2 The 'seemingly unrelated' simultaneous equations

First, we consider the model under the restrictions $\beta^*_{12} = \lambda^*_{12} = \lambda^*_{21} = 0$. In terms of parameters in (18) and (19), we have the restrictions $\beta_{12} = \beta_{21} = \lambda_{12} = \lambda_{21} = 0$, such that the econometric model

[8] In this paper, we assume $\boldsymbol{G}, \boldsymbol{X}$ are nonstochastic. This assumption can be easily relaxed and the results will be conditional on $\boldsymbol{G}, \boldsymbol{X}$.

(22) and (23) reduces to (see (10) and (11)):

$$
\begin{aligned}
\boldsymbol{y}_1 &= \lambda_{11}\boldsymbol{G}\boldsymbol{y}_1 + \boldsymbol{X}\boldsymbol{c}_1 + \boldsymbol{G}\boldsymbol{X}\boldsymbol{\gamma}_1 + \boldsymbol{\epsilon}_1 &(24)\\
\boldsymbol{y}_2 &= \lambda_{22}\boldsymbol{G}\boldsymbol{y}_2 + \boldsymbol{X}\boldsymbol{c}_2 + \boldsymbol{G}\boldsymbol{X}\boldsymbol{\gamma}_2 + \boldsymbol{\epsilon}_2. &(25)
\end{aligned}
$$

In this case, an individual's decision is still allowed to be correlated with his/her other decisions through the error term.

Without loss of generality, we focus on the identification of equation (24). The identification condition for (25) can be analogously derived. Let $\boldsymbol{Z}_1 = [\boldsymbol{G}\boldsymbol{y}_1, \boldsymbol{X}, \boldsymbol{G}\boldsymbol{X}]$. For the identification assumption to hold, $\mathrm{E}(\boldsymbol{Z}_1)$ needs to have full column rank. In the following proposition, we give a sufficient condition for this rank condition.[9]

Proposition 1 *Suppose* $\boldsymbol{\gamma}_1 + \lambda_{11}\boldsymbol{c}_1 \neq \boldsymbol{0}$. *Then* $\mathrm{E}(\boldsymbol{Z}_1)$ *of (24) has full column rank if and only if* $\boldsymbol{I}, \boldsymbol{G}, \boldsymbol{G}^2$ *are linearly independent.*

If $\boldsymbol{\gamma}_1 + \lambda_{11}\boldsymbol{c}_1 = \boldsymbol{0}$, then $\mathrm{E}(\boldsymbol{G}\boldsymbol{y}) = \boldsymbol{G}\boldsymbol{X}\boldsymbol{c}_1$. In this case, model (24) can not be identified because the matrix $\mathrm{E}(\boldsymbol{Z}_1) = [\mathrm{E}(\boldsymbol{G}\boldsymbol{y}), \boldsymbol{X}, \boldsymbol{G}\boldsymbol{X}]$ 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, when $\boldsymbol{\gamma}_1 + \lambda_{11}\boldsymbol{c}_1 = \boldsymbol{0}$, the reduced form of (24) becomes a simple regression model with neither endogenous nor contextual effects. Interdependence across individuals goes through unobservables (correlated disturbances) instead of observables. A special case of $\boldsymbol{\gamma}_1 + \lambda_{11}\boldsymbol{c}_1 = 0$ would be $\boldsymbol{c}_1 = \boldsymbol{\gamma}_1 = 0$. In this case, $\mathrm{E}(\boldsymbol{G}\boldsymbol{y}) = 0$, and the model cannot be identified as there is no relevant exogenous covariate in the model that can be used as an instrument for the endogenous effect.

If $\boldsymbol{\gamma}_1 + \lambda_{11}\boldsymbol{c}_1 \neq \boldsymbol{0}$, Bramoullé et al. (2009) have shown that the single-equation network model can be identified if intransitivities exist in a network so that $\boldsymbol{I}, \boldsymbol{G}, \boldsymbol{G}^2$ are linearly independent. In this case, $\boldsymbol{G}^2\boldsymbol{X}$ is not perfectly collinear with $\boldsymbol{X}$ and $\boldsymbol{G}\boldsymbol{X}$ and thus can be used as an instrument for the endogenous effect. For the SUR model, although the error term of (24) is allowed to be correlated with the error term of (25), the identification condition for (24) is the same as that for the single-equation network model given by Bramoullé et al. (2009). The presence of two equations with correlate errors does not inhibit identification. This is because the identification condition given by Bramoullé et al. (2009) is based on the mean of the reduced form equation, which is not affected by the correlation structure of the error terms.

[9] All proofs can be found in Appendix C.

3.3 The 'triangular' system of simultaneous equations

Next, we will consider the model under the restrictions $\beta^*_{12} = \lambda^*_{12} = 0$. In terms of parameters in (18) and (19), we have the restrictions $\beta_{12} = \beta_{21} = \lambda_{12} = 0$. This shuts down the simultaneity effect and one direction of the cross-choice peer effect. It retains the endogenous and contextual effects associated with each choice and retains the need to solve for the cross-choice peer effect in one direction. In this case, the model becomes (see (12) and (13)):

$$\boldsymbol{y}_1 = \lambda_{11}\boldsymbol{G}\boldsymbol{y}_1 + \lambda_{21}\boldsymbol{G}\boldsymbol{y}_2 + \boldsymbol{X}\boldsymbol{c}_1 + \boldsymbol{G}\boldsymbol{X}\boldsymbol{\gamma}_1 + \boldsymbol{\epsilon}_1 \tag{26}$$

$$\boldsymbol{y}_2 = \lambda_{22}\boldsymbol{G}\boldsymbol{y}_2 + \boldsymbol{X}\boldsymbol{c}_2 + \boldsymbol{G}\boldsymbol{X}\boldsymbol{\gamma}_2 + \boldsymbol{\epsilon}_2. \tag{27}$$

From Proposition 1, the unknown parameters in equation (27) can be identified if $\boldsymbol{\gamma}_2 + \lambda_{22}\boldsymbol{c}_2 \neq \boldsymbol{0}$ and $\boldsymbol{I}, \boldsymbol{G}, \boldsymbol{G}^2$ are linearly independent. We can next consider the identification condition for equation (26). Let $\boldsymbol{Z}_1 = [\boldsymbol{G}\boldsymbol{y}_1, \boldsymbol{G}\boldsymbol{y}_2, \boldsymbol{X}, \boldsymbol{G}\boldsymbol{X}]$. For the identification of equation (26), we need to find a sufficient condition for $\mathrm{E}(\boldsymbol{Z}_1)$ to have full column rank.

Proposition 2 *Suppose* $\boldsymbol{\gamma}_2 + \lambda_{22}\boldsymbol{c}_2 \neq \boldsymbol{0}$ *and*

$$(\lambda_{11} - \lambda_{22})(\boldsymbol{\gamma}_1 + \lambda_{11}\boldsymbol{c}_1) + \lambda_{21}(\boldsymbol{\gamma}_2 + \lambda_{11}\boldsymbol{c}_2) \neq \boldsymbol{0}. \tag{28}$$

Then $\mathrm{E}(\boldsymbol{Z}_1)$ *of model (26) has full column rank if and only if* $\boldsymbol{I}, \boldsymbol{G}, \boldsymbol{G}^2, \boldsymbol{G}^3$ *are linearly independent.*

A special case of (28) is of particular interest. If $\lambda_{21} = 0$ and $\lambda_{11} = \lambda_{22}$ so that the data generating process (DGP) is a pair of 'seemingly unrelated' simultaneous equations with identical endogenous effects, then, according to (28), (26) cannot be identified. Indeed, if $\lambda_{21} = 0$ in the DGP, $\mathrm{E}(\boldsymbol{Z}_1) = [\boldsymbol{G}\mathrm{E}(\boldsymbol{y}_1), \boldsymbol{G}\mathrm{E}(\boldsymbol{y}_2), \boldsymbol{X}, \boldsymbol{G}\boldsymbol{X}]$, where

$$\begin{aligned}
\boldsymbol{G}\mathrm{E}(\boldsymbol{y}_1) &= \boldsymbol{G}\boldsymbol{X}\boldsymbol{c}_1 + (\boldsymbol{I} - \lambda_{11}\boldsymbol{G})^{-1}\boldsymbol{G}^2\boldsymbol{X}(\lambda_{11}\boldsymbol{c}_1 + \boldsymbol{\gamma}_1), \\
\boldsymbol{G}\mathrm{E}(\boldsymbol{y}_2) &= \boldsymbol{G}\boldsymbol{X}\boldsymbol{c}_2 + (\boldsymbol{I} - \lambda_{22}\boldsymbol{G})^{-1}\boldsymbol{G}^2\boldsymbol{X}(\lambda_{22}\boldsymbol{c}_2 + \boldsymbol{\gamma}_2).
\end{aligned}$$

Thus, when $\lambda_{11} = \lambda_{22}$, $\mathrm{E}(\boldsymbol{Z}_1)$ does not have full column rank and the 'triangular' model cannot be identified. On the other hand, if $\lambda_{11} \neq \lambda_{22}$, $\lambda_{11}\boldsymbol{c}_1 + \boldsymbol{\gamma}_1 \neq 0$, and $\lambda_{22}\boldsymbol{c}_2 + \boldsymbol{\gamma}_2 \neq 0$, the 'triangular' model can be identified if $\boldsymbol{I}, \boldsymbol{G}, \boldsymbol{G}^2, \boldsymbol{G}^3$ are linearly independent.

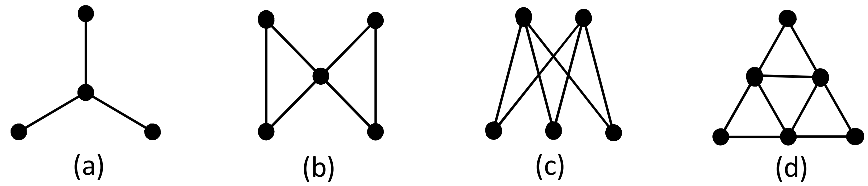


Figure 2: networks where the SUR model can be identified but the 'triangular' model cannot.

By comparing Proposition 2 with Proposition 1, we can see that the identification of the cross-choice peer effect in the 'triangular' model has higher requirement in terms of intransitivities than the SUR model. Suppose i and j are friends, j and k are friends and k and l are friends. Then, the identification of the SUR model requires that k is not a friend of i and the identification of the 'triangular' model requires that k is not a friend of i and l is not a friend of i or j. Figure 2 lists a few examples where the SUR model can be identified but the 'triangular' model cannot. It is easy to check that for all four networks in Figure 2, $\boldsymbol{I}, \boldsymbol{G}, \boldsymbol{G}^2$ are linearly independent. However, for networks (a) and (c), $\boldsymbol{G}^3 = \boldsymbol{G}$; for network (b), $\boldsymbol{G}^3 = -\frac{1}{4}\boldsymbol{I} + \frac{1}{4}\boldsymbol{G} + \boldsymbol{G}^2$; and for network (d), $\boldsymbol{G}^3 = -\frac{1}{8}\boldsymbol{I} + \frac{3}{8}\boldsymbol{G} + \frac{3}{4}\boldsymbol{G}^2$. Therefore, it follows by Propositions 1 and 2 that, for the networks in Figure 2, the SUR model can be identified but the 'triangular' model cannot.

To understand why this does not work for networks such as (a) and (c), we refer to network (a). In this star network, the center individual's friends have no other friends. The periphery individuals have some second-degree connections, but no one has a third degree connection. As a result, we cannot draw on the fact that our friends' friends may have indirect influences on our behavior.

3.4 The 'square' system of simultaneous equations

To better understand the identification challenges brought by the simultaneity effect and the cross-choice peer effect, we consider three specifications of the 'square' model. In the first case we only include the cross-choice peer effect and in the second case we only include the simultaneity effect. The third case incorporates both simultaneity and cross-choice peer effects.

3.4.1 The 'square' model with cross-choice peer effects

First, we will consider the model under the restrictions $\beta_{12}^* = 0$. In terms of parameters in (18) and (19), we have the restrictions $\beta_{12} = \beta_{21} = 0$. Thus, the model becomes (see (14) and (15)):

$$\boldsymbol{y}_1 = \lambda_{11}\boldsymbol{G}\boldsymbol{y}_1 + \lambda_{21}\boldsymbol{G}\boldsymbol{y}_2 + \boldsymbol{X}\boldsymbol{c}_1 + \boldsymbol{G}\boldsymbol{X}\boldsymbol{\gamma}_1 + \boldsymbol{\epsilon}_1 \tag{29}$$

$$\boldsymbol{y}_2 = \lambda_{22}\boldsymbol{G}\boldsymbol{y}_2 + \lambda_{12}\boldsymbol{G}\boldsymbol{y}_1 + \boldsymbol{X}\boldsymbol{c}_2 + \boldsymbol{G}\boldsymbol{X}\boldsymbol{\gamma}_2 + \boldsymbol{\epsilon}_2, \tag{30}$$

The following proposition gives a sufficient condition for the identification of (29). The sufficient condition for the identification of (30) can be analogously derived. Let $\boldsymbol{Z}_1 = [\boldsymbol{G}\boldsymbol{y}_1, \boldsymbol{G}\boldsymbol{y}_2, \boldsymbol{X}, \boldsymbol{G}\boldsymbol{X}]$ and $\rho_0, \rho_1, \rho_2, \rho_3$ be generic constant terms that may be different for different uses.

Proposition 3 *If either (i) $\boldsymbol{I}, \boldsymbol{G}, \boldsymbol{G}^2, \boldsymbol{G}^3, \boldsymbol{G}^4$ are linearly independent and $\mathbf{A}_1$ given by (35) has full rank, or (ii) $\boldsymbol{I}, \boldsymbol{G}, \boldsymbol{G}^2, \boldsymbol{G}^3$ are linearly independent, $\boldsymbol{G}^4 = \rho_0\boldsymbol{I} + \rho_1\boldsymbol{G} + \rho_2\boldsymbol{G}^2 + \rho_3\boldsymbol{G}^3$ and $\mathbf{A}_1^*$ given by (36) has full rank, then* $\mathrm{E}(\boldsymbol{Z}_1)$ *of model (29) has full column rank.*

Similarly to the 'triangular' model with a one-way cross-choice effect, Proposition 3 shows that the two-way cross-choice peer effect in a 'square' model can be identified through intransitivities that exist in a network.

3.4.2 The 'square' model with simultaneity effects

Next, we will consider the model under the restrictions $\lambda_{12}^* = \lambda_{21}^* = 0$ or $\lambda_{12} = \lambda_{21} = 0$ in terms of parameters in (18) and (19). In this case, the econometric models for the best response functions (see (16) and (17)) are:

$$\boldsymbol{y}_1 = \lambda_{11}\boldsymbol{G}\boldsymbol{y}_1 + \beta_{12}\boldsymbol{y}_2 + \boldsymbol{X}\boldsymbol{c}_1 + \boldsymbol{G}\boldsymbol{X}\boldsymbol{\gamma}_1 + \boldsymbol{\epsilon}_1 \tag{31}$$

$$\boldsymbol{y}_2 = \lambda_{22}\boldsymbol{G}\boldsymbol{y}_2 + \beta_{21}\boldsymbol{y}_1 + \boldsymbol{X}\boldsymbol{c}_2 + \boldsymbol{G}\boldsymbol{X}\boldsymbol{\gamma}_2 + \boldsymbol{\epsilon}_2, \tag{32}$$

The following proposition gives a sufficient condition for the identification of (31). The sufficient condition for the identification of (32) can be analogously derived. Let $\boldsymbol{Z}_1 = [\boldsymbol{G}\boldsymbol{y}_1, \boldsymbol{y}_2, \boldsymbol{X}, \boldsymbol{G}\boldsymbol{X}]$.

Proposition 4 *If either (i) $\boldsymbol{I}, \boldsymbol{G}, \boldsymbol{G}^2, \boldsymbol{G}^3, \boldsymbol{G}^4$ are linearly independent and $\mathbf{A}_2$ given by (37) has full rank, or (ii) $\boldsymbol{I}, \boldsymbol{G}, \boldsymbol{G}^2, \boldsymbol{G}^3$ are linearly independent, $\boldsymbol{G}^4 = \rho_0\boldsymbol{I} + \rho_1\boldsymbol{G} + \rho_2\boldsymbol{G}^2 + \rho_3\boldsymbol{G}^3$ and $\mathbf{A}_2^*$ given by (38) has full rank, then* $\mathrm{E}(\boldsymbol{Z}_1)$ *of model (31) has full column rank.*

Without social interaction effects, the simultaneity problem is a well known problem for the identification of a simultaneous equations model. The usual remedy for this identification problem is to impose exclusion restrictions on the coefficients of exogenous variables $\boldsymbol{X}$. Proposition 4 shows that, *in the absence of cross-choice peer effects, the simultaneity effect in a simultaneous equations network model can be identified without imposing exclusion restrictions on* $\boldsymbol{X}$.

3.4.3 The general 'square' model

Finally, we consider the identification of the model with both simultaneity and cross-choice peer effects. The following proposition show that, without imposing any exclusion restrictions, equations (22) and (23) cannot be identified.

Proposition 5 *The system of simultaneous equations (22) and (23) cannot be identified.*

Proposition 5 shows that, for a simultaneous equations model with both simultaneity and cross-choice peer effects, exploiting the exclusion restrictions from the intransitivities that exist in a natural network is not sufficient for the identification. One possibility to achieve identification is to impose exclusion restrictions on exogenous variables in the model. To clarify ideas, we consider the following model

$$\boldsymbol{y}_1 = \lambda_{11}\boldsymbol{G}\boldsymbol{y}_1 + \beta_{12}\boldsymbol{y}_2 + \lambda_{21}\boldsymbol{G}\boldsymbol{y}_2 + \boldsymbol{X}_1\boldsymbol{c}_1 + \boldsymbol{G}\boldsymbol{X}_1\boldsymbol{\gamma}_1 + \boldsymbol{\epsilon}_1 \quad (33)$$

$$\boldsymbol{y}_2 = \lambda_{22}\boldsymbol{G}\boldsymbol{y}_2 + \beta_{21}\boldsymbol{y}_1 + \lambda_{12}\boldsymbol{G}\boldsymbol{y}_1 + \boldsymbol{X}_2\boldsymbol{c}_2 + \boldsymbol{G}\boldsymbol{X}_2\boldsymbol{\gamma}_2 + \boldsymbol{\epsilon}_2, \quad (34)$$

where $\boldsymbol{X}_1$ and $\boldsymbol{X}_2$ are column vectors and $\boldsymbol{X}_1 \neq \boldsymbol{X}_2$. The following proposition gives a sufficient condition for the identification of (33). The sufficient condition for the identification of (34) can be analogously derived. Let $\boldsymbol{Z}_1 = [\boldsymbol{G}\boldsymbol{y}_1, \boldsymbol{y}_2, \boldsymbol{G}\boldsymbol{y}_2, \boldsymbol{X}_1, \boldsymbol{G}\boldsymbol{X}_1]$.

Proposition 6 *If (i)* $\boldsymbol{I}, \boldsymbol{G}, \boldsymbol{G}^2, \boldsymbol{G}^3, \boldsymbol{G}^4$ *are linearly independent and* $\boldsymbol{A}_3$ *given by (39) has full rank, (ii)* $\boldsymbol{I}, \boldsymbol{G}, \boldsymbol{G}^2, \boldsymbol{G}^3$ *are linearly independent,* $\boldsymbol{G}^4 = \rho_0\boldsymbol{I} + \rho_1\boldsymbol{G} + \rho_2\boldsymbol{G}^2 + \rho_3\boldsymbol{G}^3$ *and* $\boldsymbol{A}_3^*$ *given by (40) has full rank, or (iii)* $\boldsymbol{I}, \boldsymbol{G}, \boldsymbol{G}^2$ *are linearly independent,* $\boldsymbol{G}^3 = \rho_0\boldsymbol{I} + \rho_1\boldsymbol{G} + \rho_2\boldsymbol{G}^2$ *and* $\boldsymbol{A}_3^{**}$ *given by (41) has full rank, then* $\mathrm{E}(\boldsymbol{Z}_1)$ *of (33) has full column rank.*

Note that a sufficient condition for $\boldsymbol{A}_3$ in (39) to have a full rank is that c_2, γ_2 are not both zeros, $\lambda_{21} + \lambda_{11}\beta_{12} \neq 0$ and $1 - \beta_{12}\beta_{21} \neq 0$.

4 Empirical applications

4.1 Data

We use for this analysis a unique and now widely used data set provided by National Longitudinal Survey of Adolescent Health (AddHealth). The AddHealth data set provides detailed information on a large number of high school students including self-reported friendship networks.[10] This database was designed to study the impact of the social environment on adolescent behavior. The program collected national representative information on 7th-12th graders in both public and private school settings.

The survey was conducted in 1994-1995 and was designed to capture information on friends, family, school and neighborhood influences on students behaviors, including academic performance, social decisions, extracurriculars, dangerous behaviors and more. Every student attending schools on the sampling day was provided with a questionnaire that covered topics on demographics, behavioral characteristics, education, family background and critically for our purposes, friendships. This in-school survey sample included more than 90,000 students. A follow-up, more focused study covered a subset of the total sample with approximately 20,000 students. This more detailed survey included an in-home questionnaire as well as a supplemental survey with more sensitive questions. Our empirical study focus on the 20,000 students in the in-home survey study.

As illustrated above, we will mainly focus here on two activities: education (i.e. GPA) and screen activities. The former (GPA) is the average of the student most recent grades in English, science, math and history. The latter is the response to the question: "During the past week, how many times did you watch television or videos, or play video games?", coded as 0 if not at all, 1 if one or two times, 2 if three or four times, and 3 if five times or more.

The adjacency matrix $\boldsymbol{G} = [g_{ij}^*]$, where $g_{ij}^* = g_{ij} / \sum_{j=1}^{n} g_{ij}$, is constructed based on the friend-nomination information of the AddHealth data. In the questionnaire, students were asked to identify their best friends from a school roster, and then list up to five boys and five girls from this list. We assume friendship is reciprocal. Thus, for students i and j ($i \neq j$), $g_{ij} = 1$ if either i nominates j or j nominates i as a friend and $g_{ij} = 0$ otherwise. After removing students with no friends, the

[10] This research uses data from Add Health, a program project designed by J. Richard Udry, Peter S. Bearman, and Kathleen Mullan Harris, and funded by a grant P01-HD31921 from the National Institute of Child Health and Human Development, with cooperative funding from 17 other agencies. Special acknowledgment is due Ronald R. Rindfuss and Barbara Entwisle for assistance in the original design. Persons interested in obtaining data files from Add Health should contact Add Health, Carolina Population Center, 123 W. Franklin Street, Chapel Hill, NC 27516-2524 (addhealth@unc.edu). No direct support was received from grant P01-HD31921 for this analysis.

sample consists of 9,065 students distributed over 1,964 networks, with network size ranging from 2 to 1505. Because the strength of peer effect may vary with network size (see Calvó-Armengol et al., 2009), we exclude networks at the extremes of the network size distribution and focus our analysis on moderate-sized networks with network size between 10 and 200. Our selected sample consists of 1,806 students distributed over 56 networks, with network size ranging between 10 and 145. The mean and the standard deviation of network size are 32.25 and 34.23. Furthermore, in our sample, the average number of friends of a student is 3.18 with the standard deviation 2.35.

A detailed data description is available in Table 1.

[Insert Table 1 here]

4.2 The seemingly unrelated regression (SUR) case

First, we model a student's GPA and screen activities using the seemingly unrelated equations (24) and (25) with network-specific fixed effect.[11] To avoid the 'incidental parameter' problem induced by the network fixed effect dummy variables, we perform a within transformation to get rid of network dummy variables by subtracting the network average from the individual-level variables. Then, we estimate the model using Kelejian and Prucha's (2004) generalized spatial 2SLS (GS2SLS) with the IV matrix $\boldsymbol{Q} = [\boldsymbol{X}, \boldsymbol{G}\boldsymbol{X}, \boldsymbol{G}^2\boldsymbol{X}]$. The GS2SLS approach estimates the simultaneous equations model equation by equation. It is not efficient as it does not take into account the cross-equation correlation in the error term. To utilize the full system information, we also consider the generalized spatial 3SLS (GS3SLS) estimator, which jointly estimate the simultaneous equations (24) and (25). The GS3SLS estimator can be considered as a generalized least squares version of the GS2SLS estimator, which take the cross-equation correlation of the disturbances into account. The details of the estimators are given in Appendix B.3.

Table 2 reports the estimation results. For both GS2SLS and GS3SLS estimators, the well documented impact of peer GPA is observed to be quite strong. In other words, the higher is the average grades of an individual's friends, the higher is his/her own grade. Ceteris paribus, a student's GPA increases by 0.7 if the average peer GPA increases by 1 point. This result is in line with most papers showing positive peer effects in education (see, e.g. Calvó-Armengol, et al., 2009; DiGiorgi et al., 2012; Bifulco, et al., 2011). Interestingly, in this model, the time friends spend watching

[11] The details on the specification and identification of the simultaneous equations model with network fixed effects are given in Appendix D.

TV (or any screen activity) has a positive but not statistically significant impact of own time spent watching TV. This is again in line with the few empirical studies on this topic (Liu et al., 2012).

[Insert Table 2 here]

4.3 The triangular case

Table 3 shows the results of the two possible triangular cases. Triangular model 1 assumes a student's GPA is affected by the average screen activities of the peers but not vice versa. That is, in equations (26) and (27), $\boldsymbol{y}_1$ corresponds to GPA and $\boldsymbol{y}_2$ to screen activities. Triangular model 2 assumes a student's screen activities is affected by the average GPA of the peers but not vice versa. That is, in equations (26) and (27), $\boldsymbol{y}_1$ corresponds to screen activities and $\boldsymbol{y}_2$ to GPA. For the triangular models, we only report the generalized spatial estimator (GS3SLS) to save space. We notice that, for both triangular models, the estimation results of (27) are in line with those in the SUR case. Therefore, we focus on the estimation results of (26).

First, for triangular model 1 ($\boldsymbol{y}_1 =$ GPA), the estimates of λ_{11} and λ_{21} in (26) are both significant with a positive sign for the former and a negative sign for the later. This indicates that the higher the average GPA of a student's friends the higher is the grade of this student, but the more his/her friends spend time in screen activities the lower is his/her grades. A plausible explanation for the negative cross-choice peer effect (λ_{21}) is that, after accounting for the direct peer influence from studying, the more time my friends spend watching TV, the less time they spend doing homework, which leads me to spend less doing my homework too, which, in turn, decreases my grades. The endogenous peer effect and the cross-choice peer effect work in the opposite directions, but we can see that the first effect is higher in magnitude (and more significant) than the second one so that the total effect of peers on own education may still be positive.

For triangular model 2 ($\boldsymbol{y}_1 =$ screen activities), the GS3SLS estimates of λ_{11} and λ_{21} in (26) are both positive but not significant, indicating there is no statistically significant evidence that a student's screen activities depends on how much time his/her friends spend on screen activities or studying. Take video games as an example. There is, obviously, a social aspect, but there is also an addictive one. Thus, if a teenager is obsessed with video games, then the addiction itself might be the most important component to the decision and the role of social influence less salient. Therefore, it would be more difficult to find evidence of peer effect in such activities.

[Insert Table 3 here]

4.4 The square case

We consider three specifications of the square model. Square model 1 given by (29) and (30) only includes the cross-choice peer effect. Square model 2 given by (31) and (32) only includes the simultaneity effect. Square model 3 given by (33) and (34) is the most general one with both simultaneity and cross-choice peer effects. Table 4 reports the GS3SLS estimation results for square models 1 and 2. Table 5 displays the results for the GS2SLS and GS3SLS estimation of square model 3.

The square case introduce a number of new requirements for identification. Most importantly, square model 3 requires the introduction of two instruments (or exclusion restrictions), one for each decision. We instrument the GPA of a student (y_{1i}) with his/her parental education and the time spent in screen activities of a student (y_{2i}) by a variable that measures whether the respondent's parents let the student make his/her own decision on how much TV to watch (the variable "TV watching decision" in Table 1). Our first exclusion is that parental education has an impact on the student's GPA, but does not have a direct impact on the student's screen activities. Our second exclusion is that the variable "TV watching decision" has an impact on the time spent in screen activities but that the latter has no direct impact on GPA. We believe that both exclusions are reasonable and very intuitive. Furthermore, from the estimation results of the square models reported in Tables 4 and 5, we can see that parental education has a significantly positive impact on GPA and "TV watching decision" has a significantly positive impact on screen activities, which indicating the suggested instruments are both relevant. We also conduct overidentification restrictions (OIR) test for both equations, which provide empirical evidence that the suggested instruments are both valid.

For the estimation results, we focus on the case of GS3SLS, which is the more efficient estimator taking into account the cross-equation correlation in the disturbances. For the estimation of square model 1, the results are largely consistent with those found in the triangular model, which highlight the *positive peer effects in education* and the *negative cross-choice peer effect in watching TV on own education*. Interestingly, from the estimation of square model 2, the simultaneity effect is found to be positive and statistically significant for both activities. These results are confirmed from the estimation of the more general square model 3. Furthermore, estimation of square model 3 also shows positive peer effect in screen activities and negative cross-choice peer effect in GPA on screen activities. These results indicate, keeping peers' GPA and screen activities fixed and controlling for a wide range of other correlates, watching more TV could be beneficial to a student's grade. On the other hand, a student will watch more TV if his/her friends watch more TV or if his/her own GPA improves. But a student will watch less TV as his/her friends get better grades. A plausible explanation for the negative cross-choice peer effect and the positive simultaneity effect is that a student watches less TV if he/she feels pressure from peers' better academic performance, but a student may want to reward him/herself with some entertainment if his/her own grades improves.

[Insert Tables 4 and 5 here]

5 Conclusion

In this paper, we investigate the impact of peers on own outcomes where all agents embedded in a network choose more than one activity. We develop a simple network model that illustrates these issues. We differentiate between the 'seemingly unrelated' simultaneous equations model where people are influenced only by others within the same activity, the 'triangular' system of simultaneous equations model, where there is some asymmetry in the peers' cross effects, and the 'square' system of simultaneous equations model, where all possible cross-effects are taken into account. We develop the conditions under which each model is identified, showing that the standard network conditions for one activity are not enough for the last model to be identified. We then study the impact of peer effects on education and watching TV (and playing video games) and show that the estimation of two group-level effects has a non-trivial impact on adolescent behavior.

We believe that the methodology developed in this paper is important because, in real-world situations, individuals choose more than one activity. This also has important policy implications because some activities reinforce each other (when they are complements, like for example smoking and drinking) while other counter each other (when they are substitutes, like for example drinking and education). We have shown that education and screen activities have a non-trivial relationship and that, in some cases, watching TV could be good for education. In terms of policy implications, this implies that the planner will have more than one instrument available. Consider, for example, peer effects in crime (Patacchini and Zenou, 2012). Most policies aiming at reducing crime will either punish more or, in the context of criminal networks, target some "key players", who once removed generates the highest possible reduction in crime (Liu et al., 2012). Imagine, now, that using our methodology, one founds that crime and education are substitutes (which should be the case) and can *quantify* the impact of peers on each activity and see how much their offset each other. Then, the planner can identify a key player based on a two-dimensional scale so that it is the criminal who once removed reduces the most total crime *and* reduces the most total grades in a network. In other words, a planner can implement a policy that impacts on both crime and education by providing incentives to reduce crime and increase education. This is what we plan to investigate in the future.

References

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

Belhaj, M. and F. Deroïan (2012), "Competing activities in social networks," Unpublished manuscript, Aix-Marseille School of Economics.

Bifulco, R., Fletcher, J.M. and S.L. Ross (2011), "The effect of classmate characteristics on post-secondary outcomes: Evidence from the Add Health," *American Economic Journal: Economic Policy* 3, 25-53.

Blume, L.E., Brock, W.A., Durlauf, S.N. and Y.M. Ioannides (2011), "Identification of social interactions," In: J. Benhabib, A. Bisin, and M.O. Jackson (Eds.), *Handbook of Social Economics, Vol. 1B*, Amsterdam: Elsevier Publisher, pp. 853-964.

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

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

Bramoullé, Y., Kranton, R.E. and M. D'Amours (2012), "Strategic interaction and networks," Unpublished manuscript, Duke University.

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

Cohen-Cole E. (2006), "Multiple groups identification in the linear-in-means model," *Economic Letters* 92, 157–162.

Cutler, D.M. and E.L. Glaeser (2010), "Social interactions and smoking," In: D. Wise (Ed.), *Research Findings in the Economics of Aging*, Chicago: University of Chicago Press, pp. 123-141.

De Giorgi, G., Pellizzari m. and S. Redaelli (2010), "Identification of social interactions through partially overlapping peer groups," *American Economic Journal: Applied Economics* 2, 241-275.

Durlauf, S.N. and Y.M. Ioannides (2010), "Social interactions," *Annual Review of Economics* 2, 451-478.

Fletcher, J.M. (2012), "Peer influences on adolescent alcohol consumption: Evidence using an instrumental variables/fixed effect approach," *Journal of Population Economics* 25, 1265-1286.

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

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

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

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

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

Liu, X., Patacchini, E. and Y. Zenou (2012), "Peer effects in education, sport and screen activities. Local aggregate or local average?", CEPR Discussion Paper No. 8477.

Liu, X., Patacchini, E., Zenou, Y. and L-F. Lee (2012), "Criminal networks: Who is the key player?" CEPR Discussion Paper No. 8772.

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

Sacerdote, B. (2011), "Peer effects in education: How might they work, how big are they and how much do we know thus far?", In: E.A. Hanushek, S. Machin and L. Woessmann (Eds.), *Handbook of Economics of Education*, *Vol. 3*, Amsterdam: Elevier Science, pp. 249-277.

APPENDIX

A Rank Conditions for Identification

In this appendix we list the matrices whose rank conditions are used for the identification of the 'square' model.

For the 'square' model with cross-choice peer effects, identification requires $\boldsymbol{A}_1$ or $\boldsymbol{A}_1^*$ to have full rank (see Proposition 3).

$$
\begin{aligned}
\boldsymbol{A}_1 &= [a_{1,ij}] \qquad (35)\\
&= \begin{bmatrix}
0 & 0 & 1 & 0 \\
c_1 & c_2 & -(2\lambda_{11}+\lambda_{22}) & 1 \\
\gamma_1-(\lambda_{11}+\lambda_{22})c_1+\lambda_{21}c_2 & \gamma_2+\lambda_{12}c_1-2\lambda_{11}c_2 & \lambda_{11}^2+2\lambda_{11}\lambda_{22}-\lambda_{12}\lambda_{21} & -(2\lambda_{11}+\lambda_{22}) \\
\begin{bmatrix} \lambda_{11}\lambda_{22}c_1-\lambda_{11}\lambda_{21}c_2 \\ -(\lambda_{11}+\lambda_{22})\gamma_1+\lambda_{21}\gamma_2 \end{bmatrix} & \lambda_{12}\gamma_1-2\lambda_{11}\gamma_2-\lambda_{11}\lambda_{12}c_1+\lambda_{11}^2c_2 & -\lambda_{11}(\lambda_{11}\lambda_{22}-\lambda_{12}\lambda_{21}) & \lambda_{11}^2+2\lambda_{11}\lambda_{22}-\lambda_{12}\lambda_{21} \\
\lambda_{11}(\lambda_{22}\gamma_1-\lambda_{21}\gamma_2) & -\lambda_{11}(\lambda_{12}\gamma_1-\lambda_{11}\gamma_2) & 0 & -\lambda_{11}(\lambda_{11}\lambda_{22}-\lambda_{12}\lambda_{21})
\end{bmatrix}.
\end{aligned}
$$

$$
\boldsymbol{A}_1^* = \begin{bmatrix}
a_{1,11}+\rho_0 a_{1,51} & a_{1,12}+\rho_0 a_{1,52} & a_{1,13}+\rho_0 a_{1,53} & a_{1,14}+\rho_0 a_{1,54} \\
a_{1,21}+\rho_1 a_{1,51} & a_{1,22}+\rho_1 a_{1,52} & a_{1,23}+\rho_1 a_{1,53} & a_{1,24}+\rho_1 a_{1,54} \\
a_{1,31}+\rho_2 a_{1,51} & a_{1,32}+\rho_2 a_{1,52} & a_{1,33}+\rho_2 a_{1,53} & a_{1,34}+\rho_2 a_{1,54} \\
a_{1,41}+\rho_3 a_{1,51} & a_{1,42}+\rho_3 a_{1,52} & a_{1,43}+\rho_3 a_{1,53} & a_{1,44}+\rho_3 a_{1,54}
\end{bmatrix}. \qquad (36)
$$

For the 'square' model with simultaneity effects, identification requires $\boldsymbol{A}_2$ or $\boldsymbol{A}_2^*$ to have full rank (see Proposition 4).

$$
\begin{aligned}
\boldsymbol{A}_2 &= [a_{2,ij}] \\
&= \begin{bmatrix}
0 & \beta_{21}c_1 + c_2 & 1 - \beta_{12}\beta_{21} & 0 \\
c_1 + \beta_{12}c_2 & \begin{bmatrix} \beta_{21}\gamma_1 + \gamma_2 \\ -\lambda_{11}\beta_{21}c_1 - 2\lambda_{11}c_2 \end{bmatrix} & \begin{bmatrix} -(\lambda_{11} + \lambda_{22}) \\ -\lambda_{11}(1 - \beta_{12}\beta_{21}) \end{bmatrix} & 1 - \beta_{12}\beta_{21} \\
\begin{bmatrix} \gamma_1 + \beta_{12}\gamma_2 \\ -(\lambda_{11} + \lambda_{22})c_1 - \lambda_{11}\beta_{12}c_2 \end{bmatrix} & -\lambda_{11}(\beta_{21}\gamma_1 + 2\gamma_2 - \lambda_{11}c_2) & \lambda_{11}(\lambda_{11} + 2\lambda_{22}) & \begin{bmatrix} -(\lambda_{11} + \lambda_{22}) \\ -\lambda_{11}(1 - \beta_{12}\beta_{21})] \end{bmatrix} \\
\lambda_{11}\lambda_{22}c_1 - (\lambda_{11} + \lambda_{22})\gamma_1 - \lambda_{11}\beta_{12}\gamma_2 & \lambda_{11}^2\gamma_2 & -\lambda_{11}^2\lambda_{22} & \lambda_{11}(\lambda_{11} + 2\lambda_{22}) \\
\lambda_{11}\lambda_{22}\gamma_1 & 0 & 0 & -\lambda_{11}^2\lambda_{22}
\end{bmatrix}.
\end{aligned} \tag{37}
$$

$$
\boldsymbol{A}_2^* = \begin{bmatrix}
a_{2,11} + \rho_0 a_{2,51} & a_{2,12} + \rho_0 a_{2,52} & a_{2,13} + \rho_0 a_{2,53} & a_{2,14} + \rho_0 a_{2,54} \\
a_{2,21} + \rho_1 a_{2,51} & a_{2,22} + \rho_1 a_{2,52} & a_{2,23} + \rho_1 a_{2,53} & a_{2,24} + \rho_1 a_{2,54} \\
a_{2,31} + \rho_2 a_{2,51} & a_{2,32} + \rho_2 a_{2,52} & a_{2,33} + \rho_2 a_{2,53} & a_{2,34} + \rho_2 a_{2,54} \\
a_{2,41} + \rho_3 a_{2,51} & a_{2,42} + \rho_3 a_{2,52} & a_{2,43} + \rho_3 a_{2,53} & a_{2,44} + \rho_3 a_{2,54}
\end{bmatrix}. \tag{38}
$$

For the general 'square' model, identification requires $\boldsymbol{A}_3$, $\boldsymbol{A}_3^*$ or $\boldsymbol{A}_3^{**}$ to have full rank (see Proposition 6).

$$
\boldsymbol{A}_3 = [\boldsymbol{B}', \boldsymbol{C}']', \tag{39}
$$

where

$$\boldsymbol{B} = [b_{ij}] = \begin{bmatrix} 0 & \beta_{21}c_1 & 0 & 1-\beta_{12}\beta_{21} & 0 \\ c_1d_1 & b_{22} & \beta_{21}c_1 & b_{24} & 1-\beta_{12}\beta_{21} \\ \gamma_1-(\lambda_{11}+\lambda_{22})c_1 & b_{32} & b_{33} & b_{34} & b_{35} \\ \lambda_{11}\lambda_{22}c_1-(\lambda_{11}+\lambda_{22})\gamma_1 & -\lambda_{11}\lambda_{12}\gamma_1 & b_{43} & b_{44} & b_{45} \\ \lambda_{11}\lambda_{22}\gamma_1 & 0 & -\lambda_{11}\lambda_{12}\gamma_1 & 0 & -\lambda_{11}(\lambda_{11}\lambda_{22}-\lambda_{12}\lambda_{21}) \end{bmatrix},$$

$$\begin{aligned} b_{22} &= \beta_{21}\gamma_1 + (\lambda_{12}-\lambda_{11}\beta_{21})c_1, \\ b_{24} &= -(2\lambda_{11}+\lambda_{22}+\lambda_{12}\beta_{12}+\lambda_{21}\beta_{21}-\lambda_{11}\beta_{12}\beta_{21}), \end{aligned}$$

$$\begin{aligned} b_{32} &= (\lambda_{12}-\lambda_{11}\beta_{21})\gamma_1 - \lambda_{11}\lambda_{12}c_1, \\ b_{33} &= \beta_{21}\gamma_1 + (\lambda_{12}-\lambda_{11}\beta_{21})c_1, \\ b_{34} &= \lambda_{11}^2 + 2\lambda_{11}\lambda_{22} - \lambda_{12}\lambda_{21} + \lambda_{11}\lambda_{12}\beta_{12} + \lambda_{11}\lambda_{21}\beta_{21}, \\ b_{35} &= -[\lambda_{11}+\lambda_{22}+\lambda_{12}\beta_{12}+\lambda_{21}\beta_{21}+\lambda_{11}(1-\beta_{12}\beta_{21})], \end{aligned}$$

$$\begin{aligned} b_{43} &= (\lambda_{12}-\lambda_{11}\beta_{21})\gamma_1 - \lambda_{11}\lambda_{12}c_1, \\ b_{44} &= -\lambda_{11}(\lambda_{11}\lambda_{22}-\lambda_{12}\lambda_{21}), \\ b_{45} &= \lambda_{11}^2 + 2\lambda_{11}\lambda_{22} - \lambda_{12}\lambda_{21} + \lambda_{11}\lambda_{12}\beta_{12} + \lambda_{11}\lambda_{21}\beta_{21}, \end{aligned}$$

and

$$\boldsymbol{C} = [c_{ij}] = \begin{bmatrix} 0 & c_2 & 0 & 0 & 0 \\ \beta_{12}c_2 & \gamma_2 - 2\lambda_{11}c_2 & c_2 & 0 & 0 \\ (\lambda_{21} - \lambda_{11}\beta_{12})c_2 + \beta_{12}\gamma_2 & \lambda_{11}^2 c_2 - 2\lambda_{11}\gamma_2 & \gamma_2 - 2\lambda_{11}c_2 & 0 & 0 \\ (\lambda_{21} - \lambda_{11}\beta_{12})\gamma_2 - \lambda_{11}\lambda_{21}c_2 & \lambda_{11}^2\gamma_2 & \lambda_{11}^2 c_2 - 2\lambda_{11}\gamma_2 & 0 & 0 \\ -\lambda_{11}\lambda_{21}\gamma_2 & 0 & \lambda_{11}^2\gamma_2 & 0 & 0 \end{bmatrix}.$$

$$\boldsymbol{A}_3^* = [\boldsymbol{B}^{*\prime}, \boldsymbol{C}^{*\prime}]', \tag{40}$$

where

$$\boldsymbol{B}^* = \begin{bmatrix} b_{11} + \rho_0 b_{51} & b_{12} + \rho_0 b_{52} & b_{13} + \rho_0 b_{53} & b_{14} + \rho_0 b_{54} & b_{15} + \rho_0 b_{55} \\ b_{21} + \rho_1 b_{51} & b_{22} + \rho_1 b_{52} & b_{23} + \rho_1 b_{53} & b_{24} + \rho_1 b_{54} & b_{25} + \rho_1 b_{55} \\ b_{31} + \rho_2 b_{51} & b_{32} + \rho_2 b_{52} & b_{33} + \rho_2 b_{53} & b_{34} + \rho_2 b_{54} & b_{35} + \rho_2 b_{55} \\ b_{41} + \rho_3 b_{51} & b_{42} + \rho_3 b_{52} & b_{43} + \rho_3 b_{53} & b_{44} + \rho_3 b_{54} & b_{45} + \rho_3 b_{55} \end{bmatrix},$$

and

$$\boldsymbol{C}^* = \begin{bmatrix} c_{11} + \rho_0 c_{51} & c_{12} + \rho_0 c_{52} & c_{13} + \rho_0 c_{53} & 0 & 0 \\ c_{21} + \rho_1 c_{51} & c_{22} + \rho_1 c_{52} & c_{23} + \rho_1 c_{53} & 0 & 0 \\ c_{31} + \rho_2 c_{51} & c_{32} + \rho_2 c_{52} & c_{33} + \rho_2 c_{53} & 0 & 0 \\ c_{41} + \rho_3 c_{51} & c_{42} + \rho_3 c_{52} & c_{43} + \rho_3 c_{53} & 0 & 0 \end{bmatrix}.$$

$$\boldsymbol{A}_3^{**} = [\boldsymbol{B}^{**\prime}, \boldsymbol{C}^{**\prime}]', \tag{41}$$

where

$$
\boldsymbol{B}^{**} = \begin{bmatrix}
b_{11} + \rho_0 b_{41} + \rho_0\rho_2 b_{51} & b_{12} + \rho_0 b_{42} + \rho_0\rho_2 b_{52} & b_{13} + \rho_0 b_{43} + \rho_0\rho_2 b_{53} & b_{14} + \rho_0 b_{44} + \rho_0\rho_2 b_{54} & b_{15} + \rho_0 b_{45} + \rho_0\rho_2 b_{55} \\
\begin{bmatrix} b_{21} + \rho_1 b_{41} \\ +(\rho_0 + \rho_1\rho_2) b_{51} \end{bmatrix} & \begin{bmatrix} b_{22} + \rho_1 b_{42} \\ +(\rho_0 + \rho_1\rho_2) b_{52} \end{bmatrix} & \begin{bmatrix} b_{23} + \rho_1 b_{43} \\ +(\rho_0 + \rho_1\rho_2) b_{53} \end{bmatrix} & \begin{bmatrix} b_{24} + \rho_1 b_{44} \\ +(\rho_0 + \rho_1\rho_2) b_{54} \end{bmatrix} & \begin{bmatrix} b_{25} + \rho_1 b_{45} \\ +(\rho_0 + \rho_1\rho_2) b_{55} \end{bmatrix} \\
\begin{bmatrix} b_{31} + \rho_2 b_{41} \\ +(\rho_1 + \rho_2^2) b_{51} \end{bmatrix} & \begin{bmatrix} b_{32} + \rho_2 b_{42} \\ +(\rho_1 + \rho_2^2) b_{52} \end{bmatrix} & \begin{bmatrix} b_{33} + \rho_2 b_{43} \\ +(\rho_1 + \rho_2^2) b_{53} \end{bmatrix} & \begin{bmatrix} b_{34} + \rho_2 b_{44} \\ +(\rho_1 + \rho_2^2) b_{54} \end{bmatrix} & \begin{bmatrix} b_{35} + \rho_2 b_{45} \\ +(\rho_1 + \rho_2^2) b_{55} \end{bmatrix}
\end{bmatrix},
$$

and

$$
\boldsymbol{C}^{**} = \begin{bmatrix}
c_{11} + \rho_0 c_{41} + \rho_0\rho_2 c_{51} & c_{12} + \rho_0 c_{42} + \rho_0\rho_2 c_{52} & c_{13} + \rho_0 c_{43} + \rho_0\rho_2 c_{53} & 0 & 0 \\
c_{21} + \rho_1 c_{41} + (\rho_0 + \rho_1\rho_2) c_{51} & c_{22} + \rho_1 c_{42} + (\rho_0 + \rho_1\rho_2) c_{52} & c_{23} + \rho_1 c_{43} + (\rho_0 + \rho_1\rho_2) c_{53} & 0 & 0 \\
c_{31} + \rho_2 c_{41} + (\rho_1 + \rho_2^2) c_{51} & c_{32} + \rho_2 c_{42} + (\rho_1 + \rho_2^2) c_{52} & c_{33} + \rho_2 c_{43} + (\rho_1 + \rho_2^2) c_{53} & 0 & 0
\end{bmatrix}.
$$

B A General Multivariate Choice Network Model

B.1 Multivariate choice network model without group fixed effects

Consider a general social interaction model with n agents each facing m decisions. The n agents are partitioned into $\bar{r}$ networks such that each network has n_r agents ($r = 1, \cdots, \bar{r}$) and $\sum_{r=1}^{\bar{r}} n_r = n$. Following Kelejian and Prucha (2004), for the rth network, the equilibrium decisions are captured by the following simultaneous equation model with m equations,

$$\boldsymbol{Y}_r = \boldsymbol{Y}_r\boldsymbol{B} + \boldsymbol{X}_r\boldsymbol{C} + \bar{\boldsymbol{Y}}_r\boldsymbol{\Lambda} + \bar{\boldsymbol{X}}_r\boldsymbol{\Gamma} + \boldsymbol{E}_r. \tag{42}$$

In this model, $\boldsymbol{Y}_r = (\boldsymbol{y}_{1,r}, \cdots, \boldsymbol{y}_{m,r})$, where $\boldsymbol{y}_{k,r}$ is an $n_r \times 1$ vector of cross sectional observations on the kth decision. $\boldsymbol{X}_r$ is an $n_r \times k_x$ matrix of exogenous variables. $\bar{\boldsymbol{Y}}_r = \boldsymbol{G}_r\boldsymbol{Y}_r$ and $\bar{\boldsymbol{X}}_r = \boldsymbol{G}_r\boldsymbol{X}_r$, where $\boldsymbol{G}_r$ is an $n_r \times n_r$ adjacency matrix of known constants. $\boldsymbol{E}_r = (\boldsymbol{\epsilon}_{1,r}, \cdots, \boldsymbol{\epsilon}_{m,r}) = \boldsymbol{V}_r\boldsymbol{\Sigma}^{1/2}$, where $\boldsymbol{V}_r = [v_{ki,r}]$ is an $n_r \times m$ matrix of I.I.D. innovations with zero mean and unit variance and $\boldsymbol{\Sigma}$ is a positive semi-definite symmetric $m \times m$ matrix. Finally, $\boldsymbol{B}$, $\boldsymbol{C}$, $\boldsymbol{\Lambda}$, $\boldsymbol{\Gamma}$ are correspondingly defined parameter matrices of dimension $m \times m$, $k_x \times m$, $m \times m$, $k_x \times m$, respectively.

For all $\bar{r}$ networks, it follows from (42) that

$$\boldsymbol{Y} = \boldsymbol{Y}\boldsymbol{B} + \boldsymbol{X}\boldsymbol{C} + \bar{\boldsymbol{Y}}\boldsymbol{\Lambda} + \bar{\boldsymbol{X}}\boldsymbol{\Gamma} + \boldsymbol{E}, \tag{43}$$

where $\boldsymbol{Y} = (\boldsymbol{Y}_1', \cdots, \boldsymbol{Y}_{\bar{r}}')'$, $\boldsymbol{X} = (\boldsymbol{X}_1', \cdots, \boldsymbol{X}_{\bar{r}}')'$, $\bar{\boldsymbol{Y}} = (\bar{\boldsymbol{Y}}_1', \cdots, \bar{\boldsymbol{Y}}_{\bar{r}}')'$, $\bar{\boldsymbol{X}} = (\bar{\boldsymbol{X}}_1', \cdots, \bar{\boldsymbol{X}}_{\bar{r}}')'$, and $\boldsymbol{E} = (\boldsymbol{E}_1', \cdots, \boldsymbol{E}_{\bar{r}}')'$. Let $\boldsymbol{G} = \text{diag}\{\boldsymbol{G}_r\}_{r=1}^{\bar{r}}$, then $\bar{\boldsymbol{Y}} = \boldsymbol{G}\boldsymbol{Y}$ and $\bar{\boldsymbol{X}} = \boldsymbol{G}\boldsymbol{X}$. Let $\boldsymbol{y} = \text{vec}(\boldsymbol{Y})$, $\boldsymbol{x} = \text{vec}(\boldsymbol{X})$, and $\boldsymbol{\epsilon} = \text{vec}(\boldsymbol{E})$.[12] As $\text{vec}(\bar{\boldsymbol{Y}}\boldsymbol{\Lambda}) = \text{vec}(\boldsymbol{G}\boldsymbol{Y}\boldsymbol{\Lambda}) = (\boldsymbol{\Lambda}' \otimes \boldsymbol{G})\boldsymbol{y}$, $\text{vec}(\bar{\boldsymbol{X}}\boldsymbol{\Gamma}) = \text{vec}(\boldsymbol{G}\boldsymbol{X}\boldsymbol{\Gamma}) = (\boldsymbol{\Gamma}' \otimes \boldsymbol{G})\boldsymbol{x}$, $\text{vec}(\boldsymbol{Y}\boldsymbol{B}) = (\boldsymbol{B}' \otimes \boldsymbol{I}_n)\boldsymbol{y}$ and $\text{vec}(\boldsymbol{X}\boldsymbol{C}) = (\boldsymbol{C}' \otimes \boldsymbol{I}_n)\boldsymbol{x}$, it follows from (43) that

$$\boldsymbol{y} = \boldsymbol{B}^*\boldsymbol{y} + \boldsymbol{C}^*\boldsymbol{x} + \boldsymbol{\epsilon} \tag{44}$$

where $\boldsymbol{B}^* = (\boldsymbol{B}' \otimes \boldsymbol{I}_n + \boldsymbol{\Lambda}' \otimes \boldsymbol{G})$ and $\boldsymbol{C}^* = (\boldsymbol{C}' \otimes \boldsymbol{I}_n + \boldsymbol{\Gamma}' \otimes \boldsymbol{G})$.

Let $\boldsymbol{V} = (\boldsymbol{V}_1', \cdots, \boldsymbol{V}_{\bar{r}}')'$. Note that $\boldsymbol{\epsilon} = \text{vec}(\boldsymbol{E}) = \text{vec}(\boldsymbol{V}\boldsymbol{\Sigma}^{1/2}) = (\boldsymbol{\Sigma}^{1/2} \otimes \boldsymbol{I}_n)\text{vec}(\boldsymbol{V})$. Hence

[12] For an $n \times m$ matrix $\boldsymbol{A} = [a_{ij}]$, the vectorization of $\boldsymbol{A}$ is given by

$$\text{vec}(\boldsymbol{A}) = (a_{11}, \cdots, a_{n1}, a_{12}, \cdots, a_{n2}, \cdots, a_{1m}, \cdots, a_{nm})'.$$

Note that if $\boldsymbol{A}$, $\boldsymbol{B}$, $\boldsymbol{C}$ are conformable matrices, then $\text{vec}(\boldsymbol{A}\boldsymbol{B}\boldsymbol{C}) = (\boldsymbol{C}' \otimes \boldsymbol{A})\text{vec}(\boldsymbol{B})$, where $\otimes$ denotes the Kronecker product.

$\mathrm{E}(\boldsymbol{\epsilon}) = 0$ and $\mathrm{E}(\boldsymbol{\epsilon}\boldsymbol{\epsilon}') = (\boldsymbol{\Sigma}^{1/2} \otimes \boldsymbol{I}_n)(\boldsymbol{\Sigma}^{1/2} \otimes \boldsymbol{I}_n) = \boldsymbol{\Sigma} \otimes \boldsymbol{I}_n$. Thus, this specification allows for the innovations corresponding to the same agent to be correlated across decisions.

For the model (44) to have a unique equilibrium, we need the following assumption.

Assumption 1. $(\boldsymbol{I}_{mn} - \boldsymbol{B}^*)$ is nonsingular.

From (44), we have

$$(\boldsymbol{I}_{mn} - \boldsymbol{B}^*)\boldsymbol{y} = \boldsymbol{C}^*\boldsymbol{x} + \boldsymbol{\epsilon}.$$

If $(\boldsymbol{I}_{mn} - \boldsymbol{B}^*)$ is nonsingular, the reduced form equation is

$$\boldsymbol{y} = (\boldsymbol{I}_{mn} - \boldsymbol{B}^*)^{-1}\boldsymbol{C}^*\boldsymbol{x} + (\boldsymbol{I}_{mn} - \boldsymbol{B}^*)^{-1}\boldsymbol{\epsilon}. \tag{45}$$

Finally, we impose exclusion restrictions on (43). Let $\boldsymbol{b}_k$, $\boldsymbol{c}_k$, $\boldsymbol{\lambda}_k$ and $\boldsymbol{\gamma}_k$ be the vectors of nonzero elements of the kth columns of $\boldsymbol{B}$, $\boldsymbol{C}$, $\boldsymbol{\Lambda}$, and $\boldsymbol{\Gamma}$. Let $\mathbf{Y}_k$, $\boldsymbol{X}_k$, $\bar{\mathbf{Y}}_k$ and $\bar{\boldsymbol{X}}_k$ be the corresponding matrices of $\mathbf{Y}$, $\boldsymbol{X}$, $\bar{\mathbf{Y}}$ and $\bar{\boldsymbol{X}}$ that appear in the kth equation. Then,

$$\boldsymbol{y}_k = \boldsymbol{Z}_k\boldsymbol{\delta}_k + \boldsymbol{\epsilon}_k,$$

where $\boldsymbol{y}_k = (\boldsymbol{y}'_{k,1}, \cdots, \boldsymbol{y}'_{k,\bar{r}})'$, $\boldsymbol{Z}_k = (\mathbf{Y}_k, \boldsymbol{X}_k, \bar{\mathbf{Y}}_k, \bar{\boldsymbol{X}}_k)$, $\boldsymbol{\epsilon}_k = (\boldsymbol{\epsilon}'_{k,1}, \cdots, \boldsymbol{\epsilon}'_{k,\bar{r}})'$, and $\boldsymbol{\delta}_k = (\boldsymbol{b}_k, \boldsymbol{c}_k, \boldsymbol{\lambda}_k, \boldsymbol{\gamma}_k)$. Let $\boldsymbol{Q}$ denote the (nonstochastic) IV matrix. Then, the model is identified if the following condition is satisfied.

Assumption 2. $\lim_{n\to\infty} \frac{1}{n}\boldsymbol{Q}'\mathrm{E}(\boldsymbol{Z}_k)$ is a finite matrix which has full column rank, $k = 1, \cdots, m$.

This identification assumption implies the rank condition that $\mathrm{E}(\boldsymbol{Z}_k)$ has full column rank and that $\boldsymbol{Q}$ has a rank at least as high as $\mathrm{E}(\boldsymbol{Z}_k)$, for large enough n. This identification assumption also requires $\boldsymbol{Q}$ to be correlated with $\boldsymbol{Z}_k$ in the limit as n goes to infinity.

B.2 Multivariate choice network model with group fixed effects

Now, we allow for unobserved correlation among observations in a network by adding a group-specific effect into the model such that

$$\mathbf{Y}_r = \boldsymbol{a}_r \otimes \boldsymbol{l}_{n_r} + \mathbf{Y}_r\boldsymbol{B} + \boldsymbol{X}_r\boldsymbol{C} + \bar{\mathbf{Y}}_r\boldsymbol{\Lambda} + \bar{\boldsymbol{X}}_r\boldsymbol{\Gamma} + \boldsymbol{E}_r, \tag{46}$$

where $\boldsymbol{a}_r = (a_{1,r}, \cdots, a_{m,r})$. We allow the group-specific effect $a_{k,r}$ to depend on $\boldsymbol{X}_r$ and $\boldsymbol{G}_r$ by treating $a_{k,r}$ as an unknown parameter (as in a fixed-effect panel data model). To avoid the "incidental parameters" problem, we transform model (46) to eliminate the group fixed effect $a_{k,r}$.

First, we consider the transformation with the projection matrix $\boldsymbol{J}_r = \boldsymbol{I}_{n_r} - \frac{1}{n_r}\boldsymbol{l}_{n_r}\boldsymbol{l}'_{n_r}$. As $\boldsymbol{J}_r\boldsymbol{l}_{n_r} = \boldsymbol{0}$, premultiplying (46) by $\boldsymbol{J}_r$ gives

$$\boldsymbol{J}_r\boldsymbol{Y}_r = \boldsymbol{J}_r\boldsymbol{Y}_r\boldsymbol{B} + \boldsymbol{J}_r\boldsymbol{X}_r\boldsymbol{C} + \boldsymbol{J}_r\bar{\boldsymbol{Y}}_r\boldsymbol{\Lambda} + \boldsymbol{J}_r\bar{\boldsymbol{X}}_r\boldsymbol{\Gamma} + \boldsymbol{J}_r\boldsymbol{E}_r.$$

For all $\bar{r}$ networks, we have

$$\boldsymbol{J}\boldsymbol{Y} = \boldsymbol{J}\boldsymbol{Y}\boldsymbol{B} + \boldsymbol{J}\boldsymbol{X}\boldsymbol{C} + \boldsymbol{J}\bar{\boldsymbol{Y}}\boldsymbol{\Lambda} + \boldsymbol{J}\bar{\boldsymbol{X}}\boldsymbol{\Gamma} + \boldsymbol{J}\boldsymbol{E}, \tag{47}$$

where $\boldsymbol{J} = \mathrm{diag}\{\boldsymbol{J}_r\}_{r=1}^{\bar{r}}$.

For an essentially equivalent but more effective transformation, we consider the orthonormal matrix of $\boldsymbol{J}_r$ given by $[\boldsymbol{F}_r, \boldsymbol{l}_{n_r}/\sqrt{n_r}]$, where the columns of $\boldsymbol{F}_r$ are eigenvectors of $\boldsymbol{J}_r$ corresponding to the eigenvalue one. Note that $\boldsymbol{F}'_r\boldsymbol{l}_{n_r} = \boldsymbol{0}$, $\boldsymbol{F}'_r\boldsymbol{F}_r = \boldsymbol{I}_{n_r-1}$ and $\boldsymbol{F}_r\boldsymbol{F}'_r = \boldsymbol{J}_r$. Therefore, premultiplying (46) by $\boldsymbol{F}'_r$ leads to a transformed model without the incidental parameters,

$$\boldsymbol{F}'_r\boldsymbol{Y}_r = \boldsymbol{F}'_r\boldsymbol{Y}_r\boldsymbol{B} + \boldsymbol{F}'_r\boldsymbol{X}_r\boldsymbol{C} + \boldsymbol{F}'_r\bar{\boldsymbol{Y}}_r\boldsymbol{\Lambda} + \boldsymbol{F}'_r\bar{\boldsymbol{X}}_r\boldsymbol{\Gamma} + \boldsymbol{F}'_r\boldsymbol{E}_r.$$

For all $\bar{r}$ networks, we have

$$\boldsymbol{F}'\boldsymbol{Y} = \boldsymbol{F}'\boldsymbol{Y}\boldsymbol{B} + \boldsymbol{F}'\boldsymbol{X}\boldsymbol{C} + \boldsymbol{F}'\bar{\boldsymbol{Y}}\boldsymbol{\Lambda} + \boldsymbol{F}'\bar{\boldsymbol{X}}\boldsymbol{\Gamma} + \boldsymbol{F}'\boldsymbol{E}, \tag{48}$$

where $\boldsymbol{F} = \mathrm{diag}\{\boldsymbol{F}_r\}_{r=1}^{\bar{r}}$.

In order to achieve identification, we impose some exclusion restrictions on (48). Let $\boldsymbol{b}_k$, $\boldsymbol{c}_k$, $\boldsymbol{\lambda}_k$ and $\boldsymbol{\gamma}_k$ be the vectors of nonzero elements of the kth columns of $\boldsymbol{B}$, $\boldsymbol{C}$, $\boldsymbol{\Lambda}$, and $\boldsymbol{\Gamma}$. Let $\boldsymbol{Y}_k$, $\boldsymbol{X}_k$, $\bar{\boldsymbol{Y}}_k$ and $\bar{\boldsymbol{X}}_k$ be the corresponding matrices of $\boldsymbol{Y}$, $\boldsymbol{X}$, $\bar{\boldsymbol{Y}}$ and $\bar{\boldsymbol{X}}$ that appear in the kth equation. Then,

$$\boldsymbol{y}^*_k = \boldsymbol{Z}^*_k\boldsymbol{\delta}_k + \boldsymbol{\epsilon}^*_k, \tag{49}$$

where $\boldsymbol{y}^*_k = \boldsymbol{F}'\boldsymbol{y}_k$, $\boldsymbol{Z}^*_k = \boldsymbol{F}'\boldsymbol{Z}_k = \boldsymbol{F}'(\boldsymbol{Y}_k, \boldsymbol{X}_k, \bar{\boldsymbol{Y}}_k, \bar{\boldsymbol{X}}_k)$, $\boldsymbol{\epsilon}^*_k = \boldsymbol{F}'\boldsymbol{\epsilon}_k$, and $\boldsymbol{\delta}_k = (\boldsymbol{b}_k, \boldsymbol{c}_k, \boldsymbol{\lambda}_k, \boldsymbol{\gamma}_k)$.

B.3 GS2SLS and GS3SLS estimation

The transformed model (49) can be estimated by the generalized spatial 2SLS (GS2SLS) estimator in Kelejian and Prucha (2004). Let $\boldsymbol{Q}$ be the IV matrix and $\boldsymbol{Q}^* = \boldsymbol{F}'\boldsymbol{Q}$. Then, the GS2SLS estimator for $\boldsymbol{\delta}_k$ is given by

$$\hat{\boldsymbol{\delta}}_{2sls,k} = (\hat{\boldsymbol{Z}}_k^{*\prime}\boldsymbol{Z}_k^*)^{-1}\hat{\boldsymbol{Z}}_k^{*\prime}\boldsymbol{y}_k^*,$$

where $\hat{\boldsymbol{Z}}_k^* = \boldsymbol{P}^*\boldsymbol{Z}_k^*$ and $\boldsymbol{P}^* = \boldsymbol{Q}^*(\boldsymbol{Q}^{*\prime}\boldsymbol{Q}^*)^{-1}\boldsymbol{Q}^{*\prime}$. With $\boldsymbol{J} = \mathrm{diag}\{\boldsymbol{J}_r\}_{r=1}^{\bar{r}}$, the GS2SLS estimator for $\boldsymbol{\delta}_k$ can be rewritten as

$$\hat{\boldsymbol{\delta}}_{2sls,k} = (\boldsymbol{Z}_k^{*\prime}\boldsymbol{P}^*\boldsymbol{Z}_k^*)^{-1}\boldsymbol{Z}_k^{*\prime}\boldsymbol{P}^*\boldsymbol{y}_k^* = (\boldsymbol{Z}_k'\boldsymbol{P}\boldsymbol{Z}_k)^{-1}\boldsymbol{Z}_k'\boldsymbol{P}\boldsymbol{y}_k,$$

where $\boldsymbol{P} = \boldsymbol{J}\boldsymbol{Q}(\boldsymbol{Q}'\boldsymbol{J}\boldsymbol{Q})^{-1}\boldsymbol{Q}'\boldsymbol{J}$.

The GS2SLS estimator is not efficient as it does not take into account the cross equation correlation in the innovation vector. To utilize the full system information, Kelejian and Prucha (2004) have proposed a generalized spatial 3SLS (GS3SLS) estimator. Let $\boldsymbol{y}^* = (\boldsymbol{y}_1^{*\prime}, \cdots, \boldsymbol{y}_m^{*\prime})'$, $\boldsymbol{Z}^* = \mathrm{diag}\{\boldsymbol{Z}_k^*\}_{k=1}^m$, $\boldsymbol{\epsilon}^* = (\boldsymbol{\epsilon}_1^{*\prime}, \cdots, \boldsymbol{\epsilon}_m^{*\prime})'$, and $\boldsymbol{\delta} = (\boldsymbol{\delta}_1', \cdots, \boldsymbol{\delta}_m')'$. Then, for the system of m equations,

$$\boldsymbol{y}^* = \boldsymbol{Z}^*\boldsymbol{\delta} + \boldsymbol{\epsilon}^*. \tag{50}$$

Let $\boldsymbol{Z} = \mathrm{diag}\{\boldsymbol{Z}_k\}_{k=1}^m$, $\boldsymbol{Z}^* = \mathrm{diag}\{\boldsymbol{Z}_k^*\}_{k=1}^m = (\boldsymbol{I}_m \otimes \boldsymbol{F}')\boldsymbol{Z}$, and $\hat{\boldsymbol{Z}}^* = \mathrm{diag}\{\hat{\boldsymbol{Z}}_k^*\}_{k=1}^m = (\boldsymbol{I}_m \otimes \boldsymbol{P}^*)\boldsymbol{Z}^*$. As $\mathrm{E}(\boldsymbol{\epsilon}^*\boldsymbol{\epsilon}^{*\prime}) = (\boldsymbol{I}_m \otimes \boldsymbol{F}')\mathrm{E}(\boldsymbol{\epsilon}\boldsymbol{\epsilon}')(\boldsymbol{I}_m \otimes \boldsymbol{F}) = (\boldsymbol{I}_m \otimes \boldsymbol{F}')(\boldsymbol{\Sigma} \otimes \boldsymbol{I}_n)(\boldsymbol{I}_m \otimes \boldsymbol{F}) = \boldsymbol{\Sigma} \otimes \boldsymbol{I}_{n-\bar{r}}$, the (infeasible) GS3SLS is given by

$$\begin{aligned}
\tilde{\boldsymbol{\delta}}_{3sls} &= [\hat{\boldsymbol{Z}}^{*\prime}(\boldsymbol{\Sigma}^{-1} \otimes \boldsymbol{I}_{n-\bar{r}})\boldsymbol{Z}^*]^{-1}\hat{\boldsymbol{Z}}^{*\prime}(\boldsymbol{\Sigma}^{-1} \otimes \boldsymbol{I}_{n-\bar{r}})\boldsymbol{y}^* \\
&= [\boldsymbol{Z}^{*\prime}(\boldsymbol{\Sigma}^{-1} \otimes \boldsymbol{P}^*)\boldsymbol{Z}^*]^{-1}\boldsymbol{Z}^{*\prime}(\boldsymbol{\Sigma}^{-1} \otimes \boldsymbol{P}^*)\boldsymbol{y}^* \\
&= [\boldsymbol{Z}'(\boldsymbol{\Sigma}^{-1} \otimes \boldsymbol{P})\boldsymbol{Z}]^{-1}\boldsymbol{Z}'(\boldsymbol{\Sigma}^{-1} \otimes \boldsymbol{P})\boldsymbol{y}.
\end{aligned}$$

To estimate $\boldsymbol{\Sigma}$, let $\hat{\boldsymbol{\epsilon}}_k^* = \boldsymbol{J}\boldsymbol{y}_k - \boldsymbol{J}\boldsymbol{Z}_k\hat{\boldsymbol{\delta}}_{2sls,k}$. Then, $\boldsymbol{\Sigma}$ can be estimated by $\hat{\boldsymbol{\Sigma}}$ with its (k,l)th element being $\hat{\sigma}_{kl} = \hat{\boldsymbol{\epsilon}}_k^{*\prime}\hat{\boldsymbol{\epsilon}}_l^*/(n-\bar{r})$ for $k, l = 1, \cdots, m$. Hence, the feasible GS3SLS is given by

$$\hat{\boldsymbol{\delta}}_{3sls} = [\boldsymbol{Z}'(\hat{\boldsymbol{\Sigma}}^{-1} \otimes \boldsymbol{P})\boldsymbol{Z}]^{-1}\boldsymbol{Z}'(\hat{\boldsymbol{\Sigma}}^{-1} \otimes \boldsymbol{P})\boldsymbol{y}.$$

It follows by Kelejian and Prucha (2004) that the small sample distribution of $\hat{\boldsymbol{\delta}}_{3sls}$ can be approximated as $\hat{\boldsymbol{\delta}}_{3sls} \sim N(\boldsymbol{\delta}, [\boldsymbol{Z}'(\hat{\boldsymbol{\Sigma}}^{-1} \otimes \boldsymbol{P})\boldsymbol{Z}]^{-1})$.

C Proofs of Identification Results

In the following proofs, without loss of generality, let $k_x = 1$ so that $\boldsymbol{X}$ becomes a vector.

Proof of Proposition 1. The proof follows similar arguments as in the proof of Proposition 1 of Bramoullé et al. (2009). ■

Proof of Proposition 2. $\mathrm{E}(\boldsymbol{Z}_1) = [\boldsymbol{G}\mathrm{E}(\boldsymbol{y}_1), \boldsymbol{G}\mathrm{E}(\boldsymbol{y}_2), \boldsymbol{X}, \boldsymbol{G}\boldsymbol{X}]$ has full column rank if and only if

$$\boldsymbol{G}\mathrm{E}(\boldsymbol{y}_1)d_1 + \boldsymbol{G}\mathrm{E}(\boldsymbol{y}_2)d_2 + \boldsymbol{X}d_3 + \boldsymbol{G}\boldsymbol{X}d_4 = 0 \tag{51}$$

implies $d_1 = d_2 = d_3 = d_4 = 0$. From (26), $(\boldsymbol{I} - \lambda_{11}\boldsymbol{G})\mathrm{E}(\boldsymbol{y}_1) = \lambda_{21}\boldsymbol{G}\mathrm{E}(\boldsymbol{y}_2) + \boldsymbol{X}c_1 + \boldsymbol{G}\boldsymbol{X}\gamma_1$. It follows that,

$$\begin{aligned}
&(\boldsymbol{I} - \lambda_{11}\boldsymbol{G})[\boldsymbol{G}\mathrm{E}(\boldsymbol{y}_1)d_1 + \boldsymbol{G}\mathrm{E}(\boldsymbol{y}_2)d_2 + \boldsymbol{X}d_3 + \boldsymbol{G}\boldsymbol{X}d_4] \\
= \; & (\lambda_{21}\boldsymbol{G}^2\mathrm{E}(\boldsymbol{y}_2) + \boldsymbol{G}\boldsymbol{X}c_1 + \boldsymbol{G}^2\boldsymbol{X}\gamma_1)d_1 + (\boldsymbol{I} - \lambda_{11}\boldsymbol{G})[\boldsymbol{G}\mathrm{E}(\boldsymbol{y}_2)d_2 + \boldsymbol{X}d_3 + \boldsymbol{G}\boldsymbol{X}d_4] \\
= \; & [d_2\boldsymbol{G} + (\lambda_{21}d_1 - \lambda_{11}d_2)\boldsymbol{G}^2]\mathrm{E}(\boldsymbol{y}_2) + [\boldsymbol{X}d_3 + \boldsymbol{G}\boldsymbol{X}(c_1d_1 - \lambda_{11}d_3 + d_4) + \boldsymbol{G}^2\boldsymbol{X}(\gamma_1d_1 - \lambda_{11}d_4)]
\end{aligned}$$

As $(\boldsymbol{I} - \lambda_{22}\boldsymbol{G})\mathrm{E}(\boldsymbol{y}_2) = \boldsymbol{X}c_2 + \boldsymbol{G}\boldsymbol{X}\gamma_2$ from (27), we have

$$\begin{aligned}
&(\boldsymbol{I} - \lambda_{22}\boldsymbol{G})[d_2\boldsymbol{G} + (\lambda_{21}d_1 - \lambda_{11}d_2)\boldsymbol{G}^2]\mathrm{E}(\boldsymbol{y}_2) \\
= \; & [d_2\boldsymbol{G} + (\lambda_{21}d_1 - \lambda_{11}d_2)\boldsymbol{G}^2](\boldsymbol{X}c_2 + \boldsymbol{G}\boldsymbol{X}\gamma_2) \\
= \; & \boldsymbol{G}\boldsymbol{X}c_2d_2 + \boldsymbol{G}^2\boldsymbol{X}[(\gamma_2 - \lambda_{11}c_2)d_2 + \lambda_{21}c_2d_1] + \boldsymbol{G}^3\boldsymbol{X}(\lambda_{21}\gamma_2d_1 - \lambda_{11}\gamma_2d_2),
\end{aligned}$$

and

$$\begin{aligned}
&(\boldsymbol{I} - \lambda_{22}\boldsymbol{G})[\boldsymbol{X}d_3 + \boldsymbol{G}\boldsymbol{X}(c_1d_1 - \lambda_{11}d_3 + d_4) + \boldsymbol{G}^2\boldsymbol{X}(\gamma_1d_1 - \lambda_{11}d_4)] \\
= \; & \boldsymbol{X}d_3 + \boldsymbol{G}\boldsymbol{X}[c_1d_1 - (\lambda_{11} + \lambda_{22})d_3 + d_4] \\
&+ \boldsymbol{G}^2\boldsymbol{X}[(\gamma_1 - c_1\lambda_{22})d_1 + \lambda_{11}\lambda_{22}d_3 - (\lambda_{11} + \lambda_{22})d_4] - \boldsymbol{G}^3\boldsymbol{X}(\lambda_{22}\gamma_1d_1 - \lambda_{22}\lambda_{11}d_4).
\end{aligned}$$

As $(\boldsymbol{I} - \lambda_{11}\boldsymbol{G})$ and $(\boldsymbol{I} - \lambda_{22}\boldsymbol{G})$ are nonsingular and $\boldsymbol{I}, \boldsymbol{G}, \boldsymbol{G}^2, \boldsymbol{G}^3$ are linearly independent, (51) implies $d_3 = 0$ and

$$\begin{aligned} c_1 d_1 + c_2 d_2 + d_4 &= 0 && (52) \\ (\gamma_1 - \lambda_{22}c_1 + \lambda_{21}c_2)d_1 + (\gamma_2 - \lambda_{11}c_2)d_2 - (\lambda_{11} + \lambda_{22})d_4 &= 0 && (53) \\ (\lambda_{21}\gamma_2 - \lambda_{22}\gamma_1)d_1 - \lambda_{11}\gamma_2 d_2 + \lambda_{11}\lambda_{22}d_4 &= 0. && (54) \end{aligned}$$

From (52), $d_4 = -(c_1 d_1 + c_2 d_2)$. Substitution into (53) and (54) gives

$$\begin{aligned} (\gamma_1 + \lambda_{21}c_2 + \lambda_{11}c_1)d_1 + (\gamma_2 + \lambda_{22}c_2)d_2 &= 0 && (55) \\ (\lambda_{21}\gamma_2 - \lambda_{22}\gamma_1 - \lambda_{11}\lambda_{22}c_1)d_1 - \lambda_{11}(\gamma_2 + \lambda_{22}c_2)d_2 &= 0. && (56) \end{aligned}$$

Case 1: If $\lambda_{11} = 0$, from (56), $(\lambda_{21}\gamma_2 - \lambda_{22}\gamma_1)d_1 = 0$. If $\lambda_{21}\gamma_2 - \lambda_{22}\gamma_1 \neq 0$, then $d_1 = 0$, which in turn implies $d_2 = d_4 = 0$.

Case 2: If $\lambda_{11} \neq 0$, then $\lambda_{11}\times$(55)+(56) gives

$$[(\lambda_{11} - \lambda_{22})(\gamma_1 + \lambda_{11}c_1) + \lambda_{21}(\gamma_2 + \lambda_{11}c_2)]d_1 = 0.$$

If $(\lambda_{11} - \lambda_{22})(\gamma_1 + \lambda_{11}c_1) + \lambda_{21}(\gamma_2 + \lambda_{11}c_2) \neq 0$, then $d_1 = 0$, which in turn implies $d_2 = d_4 = 0$. ■

Proof of Proposition 3. Identification requires that $\mathrm{E}(\boldsymbol{Z}_1) = [\boldsymbol{G}\mathrm{E}(\boldsymbol{y}_1), \boldsymbol{G}\mathrm{E}(\boldsymbol{y}_2), \boldsymbol{X}, \boldsymbol{G}\boldsymbol{X}]$ has a full column rank for large enough n, or, equivalently,

$$\boldsymbol{G}\mathrm{E}(\boldsymbol{y}_1)d_1 + \boldsymbol{G}\mathrm{E}(\boldsymbol{y}_2)d_2 + \boldsymbol{X}d_3 + \boldsymbol{G}\boldsymbol{X}d_4 = 0 \tag{57}$$

implies $d_1 = d_2 = d_3 = d_4 = 0$. From (29) and (30), we have

$$\begin{aligned} (\boldsymbol{I} - \lambda_{11}\boldsymbol{G})\mathrm{E}(\boldsymbol{y}_1) &= \lambda_{21}\boldsymbol{G}\mathrm{E}(\boldsymbol{y}_2) + \boldsymbol{X}c_1 + \boldsymbol{G}\boldsymbol{X}\gamma_1, && (58) \\ (\boldsymbol{I} - \lambda_{22}\boldsymbol{G})\mathrm{E}(\boldsymbol{y}_2) &= \lambda_{12}\boldsymbol{G}\mathrm{E}(\boldsymbol{y}_1) + \boldsymbol{X}c_2 + \boldsymbol{G}\boldsymbol{X}\gamma_2. && (59) \end{aligned}$$

Premultiplying of (59) by $(\boldsymbol{I}-\lambda_{11}\boldsymbol{G})$ gives

$$\begin{aligned}
& (\boldsymbol{I}-\lambda_{11}\boldsymbol{G})(\boldsymbol{I}-\lambda_{22}\boldsymbol{G})\mathrm{E}(\boldsymbol{y}_2) \\
= \;& \lambda_{12}\boldsymbol{G}(\boldsymbol{I}-\lambda_{11}\boldsymbol{G})\mathrm{E}(\boldsymbol{y}_1)+(\boldsymbol{I}-\lambda_{11}\boldsymbol{G})(\boldsymbol{X}c_2+\boldsymbol{G}\boldsymbol{X}\gamma_2) \\
= \;& \lambda_{12}\boldsymbol{G}[\lambda_{21}\boldsymbol{G}\mathrm{E}(\boldsymbol{y}_2)+\boldsymbol{X}c_1+\boldsymbol{G}\boldsymbol{X}\gamma_1]+(\boldsymbol{I}-\lambda_{11}\boldsymbol{G})(\boldsymbol{X}c_2+\boldsymbol{G}\boldsymbol{X}\gamma_2) \\
= \;& \lambda_{12}\lambda_{21}\boldsymbol{G}^2\mathrm{E}(\boldsymbol{y}_2)+\boldsymbol{X}c_2+\boldsymbol{G}\boldsymbol{X}(\gamma_2+\lambda_{12}c_1-\lambda_{11}c_2)+\boldsymbol{G}^2\boldsymbol{X}(\lambda_{12}\gamma_1-\lambda_{11}\gamma_2)
\end{aligned}$$

where the second equality follows by (58). Rearranging terms, we have

$$\begin{aligned}
& [\boldsymbol{I}-(\lambda_{11}+\lambda_{22})\boldsymbol{G}+(\lambda_{11}\lambda_{22}-\lambda_{12}\lambda_{21})\boldsymbol{G}^2]\mathrm{E}(\boldsymbol{y}_2) \\
= \;& \boldsymbol{X}c_2+\boldsymbol{G}\boldsymbol{X}(\gamma_2+\lambda_{12}c_1-\lambda_{11}c_2)+\boldsymbol{G}^2\boldsymbol{X}(\lambda_{12}\gamma_1-\lambda_{11}\gamma_2).
\end{aligned} \tag{60}$$

Premultiplying of (57) by $(\mathbf{I}-\lambda_{11}\boldsymbol{G})$, It follows by (58) that

$$\boldsymbol{G}\mathrm{E}(\boldsymbol{y}_2)d_2+\boldsymbol{G}^2\mathrm{E}(\boldsymbol{y}_2)(\lambda_{21}d_1-\lambda_{11}d_2)+\boldsymbol{X}d_3+\boldsymbol{G}\boldsymbol{X}(c_1d_1-\lambda_{11}d_3+d_4)+\boldsymbol{G}^2\boldsymbol{X}(\gamma_1d_1-\lambda_{11}d_4)=0. \tag{61}$$

Premultiplying of (61) by $[\boldsymbol{I}-(\lambda_{11}+\lambda_{22})\boldsymbol{G}+(\lambda_{11}\lambda_{22}-\lambda_{12}\lambda_{21})\boldsymbol{G}^2]$ and rearranging terms, it follows by (60) that

$$\boldsymbol{X}\eta_1+\boldsymbol{G}\boldsymbol{X}\eta_2+\boldsymbol{G}^2\boldsymbol{X}\eta_3+\boldsymbol{G}^3\boldsymbol{X}\eta_4+\boldsymbol{G}^4\boldsymbol{X}\eta_5=0, \tag{62}$$

where

$$\begin{aligned}
\eta_1 &= d_3 \\
\eta_2 &= c_1d_1+c_2d_2-(2\lambda_{11}+\lambda_{22})d_3+d_4 \\
\eta_3 &= [\gamma_1-(\lambda_{11}+\lambda_{22})c_1+\lambda_{21}c_2]d_1+(\gamma_2+\lambda_{12}c_1-2\lambda_{11}c_2)d_2 \\
& \quad +(\lambda_{11}^2+2\lambda_{11}\lambda_{22}-\lambda_{12}\lambda_{21})d_3-(2\lambda_{11}+\lambda_{22})d_4 \\
\eta_4 &= [\lambda_{11}\lambda_{22}c_1-\lambda_{11}\lambda_{21}c_2-(\lambda_{11}+\lambda_{22})\gamma_1+\lambda_{21}\gamma_2]d_1+(\lambda_{12}\gamma_1-2\lambda_{11}\gamma_2-\lambda_{11}\lambda_{12}c_1+\lambda_{11}^2c_2)d_2 \\
& \quad -\lambda_{11}(\lambda_{11}\lambda_{22}-\lambda_{12}\lambda_{21})d_3+(\lambda_{11}^2+2\lambda_{11}\lambda_{22}-\lambda_{12}\lambda_{21})d_4 \\
\eta_5 &= \lambda_{11}(\lambda_{22}\gamma_1-\lambda_{21}\gamma_2)d_1-\lambda_{11}(\lambda_{12}\gamma_1-\lambda_{11}\gamma_2)d_2-\lambda_{11}(\lambda_{11}\lambda_{22}-\lambda_{12}\lambda_{21})d_4.
\end{aligned}$$

If $\boldsymbol{I}, \boldsymbol{G}, \boldsymbol{G}^2, \boldsymbol{G}^3, \boldsymbol{G}^4$ are linearly independent, then $\eta_1 = \eta_2 = \eta_3 = \eta_4 = \eta_5 = 0$. In this case, $d_1 = d_2 = d_3 = d_4 = 0$ if the rank of $\boldsymbol{A}_2$ given by (35) is 4. On the other hand, if $\boldsymbol{I}, \boldsymbol{G}, \boldsymbol{G}^2, \boldsymbol{G}^3$ are linearly independent and $\boldsymbol{G}^4 = \rho_0 \boldsymbol{I} + \rho_1 \boldsymbol{G} + \rho_2 \boldsymbol{G}^2 + \rho_3 \boldsymbol{G}^3$, we have $\eta_1 + \rho_0 \eta_5 = \eta_2 + \rho_1 \eta_5 = \eta_3 + \rho_2 \eta_5 = \eta_4 + \rho_3 \eta_5 = 0$. Then, $d_1 = d_2 = d_3 = d_4 = 0$ if the rank of $\boldsymbol{A}_1^*$ given by (36) is 4. ■

Proof of Proposition 4. Identification requires that $\mathrm{E}(\boldsymbol{Z}_1) = [\boldsymbol{G}\mathrm{E}(\boldsymbol{y}_1), \mathrm{E}(\boldsymbol{y}_2), \boldsymbol{X}, \boldsymbol{G}\boldsymbol{X}]$ has a full column rank for large enough n, or, equivalently,

$$\boldsymbol{G}\mathrm{E}(\boldsymbol{y}_1)d_1 + \mathrm{E}(\boldsymbol{y}_2)d_2 + \boldsymbol{X}d_3 + \boldsymbol{G}\boldsymbol{X}d_4 = 0 \tag{63}$$

implies $d_1 = d_2 = d_3 = d_4 = 0$. From (31) and (32), we have

$$\begin{aligned} (\boldsymbol{I} - \lambda_{11}\boldsymbol{G})\mathrm{E}(\boldsymbol{y}_1) &= \beta_{12}\mathrm{E}(\boldsymbol{y}_2) + \boldsymbol{X}c_1 + \boldsymbol{G}\boldsymbol{X}\gamma_1, && (64) \\ (\boldsymbol{I} - \lambda_{22}\boldsymbol{G})\mathrm{E}(\boldsymbol{y}_2) &= \beta_{21}\mathrm{E}(\boldsymbol{y}_1) + \boldsymbol{X}c_2 + \boldsymbol{G}\boldsymbol{X}\gamma_2. && (65) \end{aligned}$$

Premultiplying of (65) by $(\boldsymbol{I} - \lambda_{11}\boldsymbol{G})$ gives

$$\begin{aligned} & (\boldsymbol{I} - \lambda_{11}\boldsymbol{G})(\boldsymbol{I} - \lambda_{22}\boldsymbol{G})\mathrm{E}(\boldsymbol{y}_2) \\ = \; & \beta_{21}(\boldsymbol{I} - \lambda_{11}\boldsymbol{G})\mathrm{E}(\boldsymbol{y}_1) + (\boldsymbol{I} - \lambda_{11}\boldsymbol{G})(\boldsymbol{X}c_2 + \boldsymbol{G}\boldsymbol{X}\gamma_2) \\ = \; & \beta_{21}[\beta_{12}\mathrm{E}(\boldsymbol{y}_2) + \boldsymbol{X}c_1 + \boldsymbol{G}\boldsymbol{X}\gamma_1] + (\boldsymbol{I} - \lambda_{11}\boldsymbol{G})(\boldsymbol{X}c_2 + \boldsymbol{G}\boldsymbol{X}\gamma_2) \\ = \; & \beta_{12}\beta_{21}\mathrm{E}(\boldsymbol{y}_2) + \boldsymbol{X}(\beta_{21}c_1 + c_2) + \boldsymbol{G}\boldsymbol{X}(\beta_{21}\gamma_1 + \gamma_2 - \lambda_{11}c_2) - \boldsymbol{G}^2\boldsymbol{X}\lambda_{11}\gamma_2, \end{aligned}$$

where the second equality follows by (64). Rearranging terms, we have

$$[(1-\beta_{21}\beta_{12})\boldsymbol{I} - (\lambda_{11}+\lambda_{22})\boldsymbol{G} + \lambda_{11}\lambda_{22}\boldsymbol{G}^2]\mathrm{E}(\boldsymbol{y}_2) = \boldsymbol{X}(\beta_{21}c_1+c_2) + \boldsymbol{G}\boldsymbol{X}(\beta_{21}\gamma_1+\gamma_2-\lambda_{11}c_2) - \boldsymbol{G}^2\boldsymbol{X}\lambda_{11}\gamma_2. \tag{66}$$

Premultiplying of (63) by $(\boldsymbol{I} - \lambda_{11}\boldsymbol{G})$, It follows by (64) that

$$\mathrm{E}(\boldsymbol{y}_2)d_2 + \boldsymbol{G}\mathrm{E}(\boldsymbol{y}_2)(\beta_{12}d_1 - \lambda_{11}d_2) + \boldsymbol{X}d_3 + \boldsymbol{G}\boldsymbol{X}(c_1d_1 - \lambda_{11}d_3 + d_4) + \boldsymbol{G}^2\boldsymbol{X}(\gamma_1 d_1 - \lambda_{11}d_4) = 0. \tag{67}$$

Premultiplying of (67) by $[(1-\beta_{21}\beta_{12})\boldsymbol{I} - (\lambda_{11}+\lambda_{22})\boldsymbol{G} + \lambda_{11}\lambda_{22}\boldsymbol{G}^2]$ and rearranging terms, it follows by (66) that

$$\boldsymbol{X}\eta_1 + \boldsymbol{G}\boldsymbol{X}\eta_2 + \boldsymbol{G}^2\boldsymbol{X}\eta_3 + \boldsymbol{G}^3\boldsymbol{X}\eta_4 + \boldsymbol{G}^4\boldsymbol{X}\eta_5 = 0,$$

where

$$\begin{aligned}
\eta_1 &= (\beta_{21}c_1 + c_2)d_2 + (1-\beta_{12}\beta_{21})d_3 \\
\eta_2 &= (c_1 + \beta_{12}c_2)d_1 + (\beta_{21}\gamma_1 + \gamma_2 - \lambda_{11}\beta_{21}c_1 - 2\lambda_{11}c_2)d_2 \\
&\quad -[\lambda_{11} + \lambda_{22} + \lambda_{11}(1-\beta_{12}\beta_{21})]d_3 + (1-\beta_{12}\beta_{21})d_4 \\
\eta_3 &= [\gamma_1 + \beta_{12}\gamma_2 - (\lambda_{11}+\lambda_{22})c_1 - \lambda_{11}\beta_{12}c_2]d_1 - \lambda_{11}(\beta_{21}\gamma_1 + 2\gamma_2 - \lambda_{11}c_2)d_2 \\
&\quad +\lambda_{11}(\lambda_{11} + 2\lambda_{22})d_3 - [\lambda_{11} + \lambda_{22} + \lambda_{11}(1-\beta_{12}\beta_{21})]d_4 \\
\eta_4 &= [\lambda_{11}\lambda_{22}c_1 - (\lambda_{11}+\lambda_{22})\gamma_1 - \lambda_{11}\beta_{12}\gamma_2]d_1 + \lambda_{11}^2\gamma_2 d_2 - \lambda_{11}^2\lambda_{22}d_3 + \lambda_{11}(\lambda_{11}+2\lambda_{22})d_4 \\
\eta_5 &= \lambda_{11}\lambda_{22}\gamma_1 d_1 - \lambda_{11}^2\lambda_{22}d_4.
\end{aligned}$$

If $\boldsymbol{I}, \boldsymbol{G}, \boldsymbol{G}^2, \boldsymbol{G}^3, \boldsymbol{G}^4$ are linearly independent, then $\eta_1 = \eta_2 = \eta_3 = \eta_4 = \eta_5 = 0$. In this case, $d_1 = d_2 = d_3 = d_4 = 0$ if the rank of $\boldsymbol{A}_2$ given by (37) is 4. On the other hand, if $\boldsymbol{I}, \boldsymbol{G}, \boldsymbol{G}^2, \boldsymbol{G}^3$ are linearly independent and $\boldsymbol{G}^4 = \rho_0\boldsymbol{I} + \rho_1\boldsymbol{G} + \rho_2\boldsymbol{G}^2 + \rho_3\boldsymbol{G}^3$, we have $\eta_1 + \rho_0\eta_5 = \eta_2 + \rho_1\eta_5 = \eta_3 + \rho_2\eta_5 = \eta_4 + \rho_3\eta_5 = 0$. Then, $d_1 = d_2 = d_3 = d_4 = 0$ if the rank of $\boldsymbol{A}_2^*$ given by (38) is 4. ■

Proof of Propositions 5 and 6. Identification requires that $\mathrm{E}(\boldsymbol{Z}_1) = [\boldsymbol{G}\mathrm{E}(\boldsymbol{y}_1), \mathrm{E}(\boldsymbol{y}_2), \boldsymbol{G}\mathrm{E}(\boldsymbol{y}_2), \boldsymbol{X}_1, \boldsymbol{G}\boldsymbol{X}_1]$ has a full column rank for large enough n, or, equivalently,

$$\boldsymbol{G}\mathrm{E}(\boldsymbol{y}_1)d_1 + \mathrm{E}(\boldsymbol{y}_2)d_2 + \boldsymbol{G}\mathrm{E}(\boldsymbol{y}_2)d_3 + \boldsymbol{X}_1 d_4 + \boldsymbol{G}\boldsymbol{X}_1 d_5 = 0 \tag{68}$$

implies $d_1 = d_2 = d_3 = d_4 = d_5 = 0$. From (33) and (34), we have

$$\begin{aligned}
(\boldsymbol{I} - \lambda_{11}\boldsymbol{G})\mathrm{E}(\boldsymbol{y}_1) &= (\beta_{12}\boldsymbol{I} + \lambda_{21}\boldsymbol{G})\mathrm{E}(\boldsymbol{y}_2) + \boldsymbol{X}_1 c_1 + \boldsymbol{G}\boldsymbol{X}_1\gamma_1, && (69) \\
(\boldsymbol{I} - \lambda_{22}\boldsymbol{G})\mathrm{E}(\boldsymbol{y}_2) &= (\beta_{21}\boldsymbol{I} + \lambda_{12}\boldsymbol{G})\mathrm{E}(\boldsymbol{y}_1) + \boldsymbol{X}_2 c_2 + \boldsymbol{G}\boldsymbol{X}_2\gamma_2. && (70)
\end{aligned}$$

Premultiplying of (70) by $(\boldsymbol{I} - \lambda_{11}\boldsymbol{G})$ gives

$$
\begin{aligned}
& (\boldsymbol{I} - \lambda_{11}\boldsymbol{G})(\boldsymbol{I} - \lambda_{22}\boldsymbol{G})\mathrm{E}(\boldsymbol{y}_2) \\
= \ & (\beta_{21}\boldsymbol{I} + \lambda_{12}\boldsymbol{G})(\boldsymbol{I} - \lambda_{11}\boldsymbol{G})\mathrm{E}(\boldsymbol{y}_1) + (\boldsymbol{I} - \lambda_{11}\boldsymbol{G})(\boldsymbol{X}_2 c_2 + \boldsymbol{G}\boldsymbol{X}_2\gamma_2) \\
= \ & (\beta_{21}\boldsymbol{I} + \lambda_{12}\boldsymbol{G})[(\beta_{12}\boldsymbol{I} + \lambda_{21}\boldsymbol{G})\mathrm{E}(\boldsymbol{y}_2) + \boldsymbol{X}_1 c_1 + \boldsymbol{G}\boldsymbol{X}_1\gamma_1] + (\boldsymbol{I} - \lambda_{11}\boldsymbol{G})(\boldsymbol{X}_2 c_2 + \boldsymbol{G}\boldsymbol{X}_2\gamma_2) \\
= \ & (\beta_{21}\boldsymbol{I} + \lambda_{12}\boldsymbol{G})(\beta_{12}\boldsymbol{I} + \lambda_{21}\boldsymbol{G})\mathrm{E}(\boldsymbol{y}_2) \\
& + \boldsymbol{X}_1\beta_{21}c_1 + \boldsymbol{G}\boldsymbol{X}_1(\beta_{21}\gamma_1 + \lambda_{12}c_1) + \boldsymbol{G}^2\boldsymbol{X}_1\lambda_{12}\gamma_1 + \boldsymbol{X}_2 c_2 + \boldsymbol{G}\boldsymbol{X}_2(\gamma_2 - \lambda_{11}c_2) - \boldsymbol{G}^2\boldsymbol{X}_2\lambda_{11}\gamma_2
\end{aligned}
$$

where the second equality follows by (69). Rearranging terms, we have

$$
\begin{aligned}
& [(\boldsymbol{I} - \lambda_{11}\boldsymbol{G})(\boldsymbol{I} - \lambda_{22}\boldsymbol{G}) - (\beta_{21}\boldsymbol{I} + \lambda_{12}\boldsymbol{G})(\beta_{12}\boldsymbol{I} + \lambda_{21}\boldsymbol{G})]\mathrm{E}(\boldsymbol{y}_2) \\
= \ & \boldsymbol{X}_1\beta_{21}c_1 + \boldsymbol{G}\boldsymbol{X}_1(\beta_{21}\gamma_1 + \lambda_{12}c_1) + \boldsymbol{G}^2\boldsymbol{X}_1\lambda_{12}\gamma_1 + \boldsymbol{X}_2 c_2 + \boldsymbol{G}\boldsymbol{X}_2(\gamma_2 - \lambda_{11}c_2) - \boldsymbol{G}^2\boldsymbol{X}_2\lambda_{11}\gamma_2.
\end{aligned} \tag{71}
$$

Premultiplying of (68) by $(\boldsymbol{I} - \lambda_{11}\boldsymbol{G})$, It follows by (69) that

$$
\boldsymbol{G}[(\beta_{12}\boldsymbol{I} + \lambda_{21}\boldsymbol{G})\mathrm{E}(\boldsymbol{y}_2) + \boldsymbol{X}_1 c_1 + \boldsymbol{G}\boldsymbol{X}_1\gamma_1]d_1 + (\boldsymbol{I} - \lambda_{11}\boldsymbol{G})[\mathrm{E}(\boldsymbol{y}_2)d_2 + \boldsymbol{G}\mathrm{E}(\boldsymbol{y}_2)d_3 + \boldsymbol{X}_1 d_4 + \boldsymbol{G}\boldsymbol{X}_1 d_5] = 0 \tag{72}
$$

Premultiplying of (72) by $[(\boldsymbol{I} - \lambda_{11}\boldsymbol{G})(\boldsymbol{I} - \lambda_{22}\boldsymbol{G}) - (\beta_{21}\boldsymbol{I} + \lambda_{12}\boldsymbol{G})(\beta_{12}\boldsymbol{I} + \lambda_{21}\boldsymbol{G})]$ and rearranging terms, it follows by (71) that

$$
\begin{aligned}
& \boldsymbol{X}_1\eta_1 + \boldsymbol{G}\boldsymbol{X}_1\eta_2 + \boldsymbol{G}^2\boldsymbol{X}_1\eta_3 + \boldsymbol{G}^3\boldsymbol{X}_1\eta_4 + \boldsymbol{G}^4\boldsymbol{X}_1\eta_5 \\
& + \boldsymbol{X}_2\xi_1 + \boldsymbol{G}\boldsymbol{X}_2\xi_2 + \boldsymbol{G}^2\boldsymbol{X}_2\xi_3 + \boldsymbol{G}^3\boldsymbol{X}_2\xi_4 + \boldsymbol{G}^4\boldsymbol{X}_2\xi_5 = 0,
\end{aligned} \tag{73}
$$

where

$$
\begin{aligned}
\eta_1 &= \beta_{21} c_1 d_2 + (1-\beta_{12}\beta_{21}) d_4 \\
\eta_2 &= c_1 d_1 + (\beta_{21}\gamma_1 + \lambda_{12} c_1 - \lambda_{11}\beta_{21} c_1) d_2 + \beta_{21} c_1 d_3 \\
&\quad -(2\lambda_{11} + \lambda_{22} + \lambda_{12}\beta_{12} + \lambda_{21}\beta_{21} - \lambda_{11}\beta_{12}\beta_{21}) d_4 + (1-\beta_{12}\beta_{21}) d_5 \\
\eta_3 &= [\gamma_1 - (\lambda_{11}+\lambda_{22}) c_1] d_1 + (\lambda_{12}\gamma_1 - \lambda_{11}\beta_{21}\gamma_1 - \lambda_{11}\lambda_{12} c_1) d_2 \\
&\quad +(\beta_{21}\gamma_1 + \lambda_{12} c_1 - \lambda_{11}\beta_{21} c_1) d_3 + (\lambda_{11}^2 + 2\lambda_{11}\lambda_{22} - \lambda_{12}\lambda_{21} + \lambda_{11}\lambda_{12}\beta_{12} + \lambda_{11}\lambda_{21}\beta_{21}) d_4 \\
&\quad -[\lambda_{11} + \lambda_{22} + \lambda_{12}\beta_{12} + \lambda_{21}\beta_{21} + \lambda_{11}(1-\beta_{12}\beta_{21})] d_5 \\
\eta_4 &= [\lambda_{11}\lambda_{22} c_1 - (\lambda_{11}+\lambda_{22})\gamma_1] d_1 - \lambda_{11}\lambda_{12}\gamma_1 d_2 + (\lambda_{12}\gamma_1 - \lambda_{11}\beta_{21}\gamma_1 - \lambda_{11}\lambda_{12} c_1) d_3 \\
&\quad -\lambda_{11}(\lambda_{11}\lambda_{22} - \lambda_{12}\lambda_{21}) d_4 + (\lambda_{11}^2 + 2\lambda_{11}\lambda_{22} - \lambda_{12}\lambda_{21} + \lambda_{11}\lambda_{12}\beta_{12} + \lambda_{11}\lambda_{21}\beta_{21}) d_5 \\
\eta_5 &= \lambda_{11}\lambda_{22}\gamma_1 d_1 - \lambda_{11}\lambda_{12}\gamma_1 d_3 - \lambda_{11}(\lambda_{11}\lambda_{22} - \lambda_{12}\lambda_{21}) d_5
\end{aligned}
$$

and

$$
\begin{aligned}
\xi_1 &= c_2 d_2 \\
\xi_2 &= \beta_{12} c_2 d_1 + (\gamma_2 - 2\lambda_{11} c_2) d_2 + c_2 d_3 \\
\xi_3 &= (\lambda_{21} c_2 + \beta_{12}\gamma_2 - \lambda_{11}\beta_{12} c_2) d_1 + (\lambda_{11}^2 c_2 - 2\lambda_{11}\gamma_2) d_2 + (\gamma_2 - 2\lambda_{11} c_2) d_3 \\
\xi_4 &= (\lambda_{21}\gamma_2 - \lambda_{11}\lambda_{21} c_2 - \lambda_{11}\beta_{12}\gamma_2) d_1 + \lambda_{11}^2\gamma_2 d_2 + (\lambda_{11}^2 c_2 - 2\lambda_{11}\gamma_2) d_3 \\
\xi_5 &= -\lambda_{11}\lambda_{21}\gamma_2 d_1 + \lambda_{11}^2\gamma_2 d_3.
\end{aligned}
$$

If $\boldsymbol{X}_1 = \boldsymbol{X}_2$, then linear independence of $\boldsymbol{I}, \boldsymbol{G}, \boldsymbol{G}^2, \boldsymbol{G}^3, \boldsymbol{G}^4$ implies $\eta_1 + \xi_1 = \eta_2 + \xi_2 = \eta_3 + \xi_3 = \eta_4 + \xi_4 = \eta_5 + \xi_5 = 0$, which in turn implies $d_1 = (\lambda_{11}\beta_{21} + \lambda_{12}) d_2/(\beta_{12}\beta_{21} - 1)$, $d_3 = (\lambda_{21}\beta_{21} + \lambda_{22}) d_2/(\beta_{12}\beta_{21} - 1)$, $d_4 = (\beta_{21} c_1 + c_2) d_2/(\beta_{12}\beta_{21} - 1)$, and $d_5 = (\beta_{21}\gamma_1 + \gamma_2) d_2/(\beta_{12}\beta_{21} - 1)$. Thus, the model fails to be identified.

If $\boldsymbol{X}_1 \neq \boldsymbol{X}_2$ we consider three cases:

Case 1, if $\boldsymbol{I}, \boldsymbol{G}, \boldsymbol{G}^2, \boldsymbol{G}^3, \boldsymbol{G}^4$ are linearly independent, then $\eta_1 = \eta_2 = \eta_3 = \eta_4 = \eta_5 = 0$ and $\xi_1 = \xi_2 = \xi_3 = \xi_4 = \xi_5 = 0$. In this case, $d_1 = d_2 = d_3 = d_4 = 0$ if the rank of $\boldsymbol{A}_3$ given by (39) is 5.

Case 2, if $\boldsymbol{I}, \boldsymbol{G}, \boldsymbol{G}^2, \boldsymbol{G}^3$ are linearly independent and $\boldsymbol{G}^4 = \rho_0 \boldsymbol{I} + \rho_1 \boldsymbol{G} + \rho_2 \boldsymbol{G}^2 + \rho_3 \boldsymbol{G}^3$, then

$$\eta_1 + \rho_0\eta_5 = \eta_2 + \rho_1\eta_5 = \eta_3 + \rho_2\eta_5 = \eta_4 + \rho_3\eta_5 = 0$$

and

$$\xi_1 + \rho_0 \xi_5 = \xi_2 + \rho_1 \xi_5 = \xi_3 + \rho_2 \xi_5 = \xi_4 + \rho_3 \xi_5 = 0.$$

In this case, $d_1 = d_2 = d_3 = d_4 = 0$ if the rank of $\boldsymbol{A}_3^*$ given by (40) is 5.

Case 3, if $\boldsymbol{I}, \boldsymbol{G}, \boldsymbol{G}^2$ are linearly independent and $\boldsymbol{G}^3 = \rho_0 \boldsymbol{I} + \rho_1 \boldsymbol{G} + \rho_2 \boldsymbol{G}^2$, then

$$\eta_1 + \rho_0 \eta_4 + \rho_0 \rho_2 \eta_5 = \eta_2 + \rho_1 \eta_4 + (\rho_0 + \rho_1 \rho_2) \eta_5 = \eta_3 + \rho_2 \eta_4 + (\rho_1 + \rho_2^2) \eta_5 = 0$$

and

$$\xi_1 + \rho_0 \xi_4 + \rho_0 \rho_2 \xi_5 = \xi_2 + \rho_1 \xi_4 + (\rho_0 + \rho_1 \rho_2) \xi_5 = \xi_3 + \rho_2 \xi_4 + (\rho_1 + \rho_2^2) \xi_5 = 0.$$

In this case, $d_1 = d_2 = d_3 = d_4 = 0$ if the rank of $\boldsymbol{A}_3^{**}$ given by (41) is 5. ■

D Econometric Model with Group Fixed Effects

Now, we can generalize the model by introducing a group-specific effect $a_{k,r}$ $(k = 1, 2)$ so that (20) and (21) becomes

$$\boldsymbol{y}_{1,r} = \lambda_{11} \boldsymbol{G}_r \boldsymbol{y}_{1,r} + \beta_{12} \boldsymbol{y}_{2,r} + \lambda_{21} \boldsymbol{G}_r \boldsymbol{y}_{2,r} + \boldsymbol{X}_r \boldsymbol{c}_1 + \boldsymbol{G}_r \boldsymbol{X}_r \gamma_1 + a_{1,r} \boldsymbol{l}_{n_r} + \boldsymbol{\epsilon}_{1,r} \tag{74}$$

$$\boldsymbol{y}_{2,r} = \lambda_{22} \boldsymbol{G}_r \boldsymbol{y}_{2,r} + \beta_{21} \boldsymbol{y}_{1,r} + \lambda_{12} \boldsymbol{G}_r \boldsymbol{y}_{1,r} + \boldsymbol{X}_r \boldsymbol{c}_2 + \boldsymbol{G}_r \boldsymbol{X}_r \gamma_2 + a_{2,r} \boldsymbol{l}_{n_r} + \boldsymbol{\epsilon}_{2,r}. \tag{75}$$

We allow the group-specific effect $a_{k,r}$ to depend on $\boldsymbol{X}_r$ and $\boldsymbol{G}_r$ by treating $a_{k,r}$ as an unknown parameter (as in a fixed-effect panel data model). To avoid the "incidental parameters" problem, we transform the model to eliminate the group fixed effect $a_{k,r}$ and then discuss the identification of the transformed model.

To eliminate the group fixed effect $a_{k,r}$, we consider the transformation with the projection matrix $\boldsymbol{J}_r = \boldsymbol{I}_{n_r} - \frac{1}{n_r} \boldsymbol{l}_{n_r} \boldsymbol{l}'_{n_r}$. As $\boldsymbol{J}_r \boldsymbol{l}_{n_r} = \boldsymbol{0}$, premultiplying of (74) and (75) by $\boldsymbol{J}_r$ gives

$$\begin{aligned} \boldsymbol{J}_r \boldsymbol{y}_{1,r} &= \lambda_{11} \boldsymbol{J}_r \boldsymbol{G}_r \boldsymbol{y}_{1,r} + \beta_{12} \boldsymbol{J}_r \boldsymbol{y}_{2,r} + \lambda_{21} \boldsymbol{J}_r \boldsymbol{G}_r \boldsymbol{y}_{2,r} + \boldsymbol{J}_r \boldsymbol{X}_r \boldsymbol{c}_1 + \boldsymbol{J}_r \boldsymbol{G}_r \boldsymbol{X}_r \gamma_1 + \boldsymbol{J}_r \boldsymbol{\epsilon}_{1,r} \\ \boldsymbol{J}_r \boldsymbol{y}_{2,r} &= \lambda_{22} \boldsymbol{J}_r \boldsymbol{G}_r \boldsymbol{y}_{2,r} + \beta_{21} \boldsymbol{J}_r \boldsymbol{y}_{1,r} + \lambda_{12} \boldsymbol{J}_r \boldsymbol{G}_r \boldsymbol{y}_{1,r} + \boldsymbol{J}_r \boldsymbol{X}_r \boldsymbol{c}_2 + \boldsymbol{J}_r \boldsymbol{G}_r \boldsymbol{X}_r \gamma_2 + \boldsymbol{J}_r \boldsymbol{\epsilon}_{2,r}. \end{aligned}$$

For all $\bar{r}$ networks, we have

$$\boldsymbol{J}\boldsymbol{y}_1 = \lambda_{11}\boldsymbol{J}\boldsymbol{G}\boldsymbol{y}_1 + \beta_{12}\boldsymbol{J}\boldsymbol{y}_2 + \lambda_{21}\boldsymbol{J}\boldsymbol{G}\boldsymbol{y}_2 + \boldsymbol{J}\boldsymbol{X}\boldsymbol{c}_1 + \boldsymbol{J}\boldsymbol{G}\boldsymbol{X}\boldsymbol{\gamma}_1 + \boldsymbol{J}\boldsymbol{\epsilon}_1 \tag{76}$$

$$\boldsymbol{J}\boldsymbol{y}_2 = \lambda_{22}\boldsymbol{J}\boldsymbol{G}\boldsymbol{y}_2 + \beta_{21}\boldsymbol{J}\boldsymbol{y}_1 + \lambda_{12}\boldsymbol{J}\boldsymbol{G}\boldsymbol{y}_1 + \boldsymbol{J}\boldsymbol{X}\boldsymbol{c}_2 + \boldsymbol{J}\boldsymbol{G}\boldsymbol{X}\boldsymbol{\gamma}_2 + \boldsymbol{J}\boldsymbol{\epsilon}_2. \tag{77}$$

D.1 The 'seemingly unrelated' simultaneous equations

First, we consider the model under the restrictions $\beta_{12} = \beta_{21} = \lambda_{12} = \lambda_{21} = 0$, such that (76) and (77) reduce to

$$\boldsymbol{J}\boldsymbol{y}_1 = \lambda_{11}\boldsymbol{J}\boldsymbol{G}\boldsymbol{y}_1 + \boldsymbol{J}\boldsymbol{X}\boldsymbol{c}_1 + \boldsymbol{J}\boldsymbol{G}\boldsymbol{X}\boldsymbol{\gamma}_1 + \boldsymbol{J}\boldsymbol{\epsilon}_1 \tag{78}$$

$$\boldsymbol{J}\boldsymbol{y}_2 = \lambda_{22}\boldsymbol{J}\boldsymbol{G}\boldsymbol{y}_2 + \boldsymbol{J}\boldsymbol{X}\boldsymbol{c}_2 + \boldsymbol{J}\boldsymbol{G}\boldsymbol{X}\boldsymbol{\gamma}_2 + \boldsymbol{J}\boldsymbol{\epsilon}_2. \tag{79}$$

Now, consider the identification of (78). From identification assumption, to achieve identification, we need $\mathrm{E}(\boldsymbol{J}\boldsymbol{Z}_1) = \boldsymbol{J}[\boldsymbol{G}\mathrm{E}(\boldsymbol{y}_1), \boldsymbol{X}, \boldsymbol{G}\boldsymbol{X}]$ to have full column rank.

Proposition 7 *If $\boldsymbol{\gamma}_1 + \lambda_{11}\boldsymbol{c}_1 \neq \boldsymbol{0}$ and $\boldsymbol{I}, \boldsymbol{G}_r, \boldsymbol{G}_r^2, \boldsymbol{G}_r^3$ are linearly independent for some r, then $\mathrm{E}(\boldsymbol{J}\boldsymbol{Z}_1)$ of (78) has full column rank.*

Proof. $\mathrm{E}(\boldsymbol{J}\boldsymbol{Z}_1)$ has full column rank if and only if

$$\boldsymbol{J}_r[\boldsymbol{G}_r\mathrm{E}(\boldsymbol{y}_{1,r})d_1 + \boldsymbol{X}_r d_2 + \boldsymbol{G}_r\boldsymbol{X}_r d_3] = 0 \tag{80}$$

implies $d_1 = d_2 = d_3 = 0$ for some r. As $\boldsymbol{G}_r\boldsymbol{l}_{n_r} = \boldsymbol{l}_{n_r}$, it is easy to see $\boldsymbol{J}_r\boldsymbol{G}_r = \boldsymbol{J}_r\boldsymbol{G}_r\boldsymbol{J}_r$, $\boldsymbol{J}_r(\boldsymbol{I} - \lambda_{11}\boldsymbol{G}_r)^{-1} = \boldsymbol{J}_r\ (\boldsymbol{I} - \lambda_{11}\boldsymbol{G}_r)^{-1}\boldsymbol{J}_r$ and $\boldsymbol{J}_r\boldsymbol{G}_r(\boldsymbol{I} - \lambda_{11}\boldsymbol{G}_r)^{-1}\boldsymbol{l}_{n_r} = \boldsymbol{0}$ by Lemma C.1 of Lee et al. (2010). Thus, substitution of the mean of the reduced form $\mathrm{E}(\boldsymbol{y}_{1,r}) = (\boldsymbol{I} - \lambda_{11}\boldsymbol{G}_r)^{-1}(a_{1,r}\boldsymbol{l}_{n_r} + \boldsymbol{X}_r c_1 + \boldsymbol{G}_r\boldsymbol{X}_r\gamma_1)$ into (80) gives

$$\begin{aligned}
\boldsymbol{J}_r[\boldsymbol{G}_r(\boldsymbol{I} - \lambda_{11}\boldsymbol{G}_r)^{-1}(a_{1,r}\boldsymbol{l}_{n_r} + \boldsymbol{X}_r c_1 + \boldsymbol{G}_r\boldsymbol{X}_r\gamma_1)d_1 + \boldsymbol{X}_r d_2 + \boldsymbol{G}_r\boldsymbol{X}_r d_3] &= 0 \\
\boldsymbol{J}_r(\boldsymbol{I} - \lambda_{11}\boldsymbol{G}_r)^{-1}\boldsymbol{J}_r[\boldsymbol{X}_r d_2 + \boldsymbol{G}_r\boldsymbol{X}_r(c_1 d_1 - \lambda_{11}d_2 + d_3) + \boldsymbol{G}_r^2\boldsymbol{X}_r(\gamma_1 d_1 - \lambda_{11}d_3)] &= 0
\end{aligned}$$

which implies $\boldsymbol{X}_r\mu_0 + \boldsymbol{G}_r\boldsymbol{X}_r\mu_1 + \boldsymbol{G}_r^2\boldsymbol{X}_r\mu_2 = \rho\boldsymbol{l}_{n_r}$ for $\mu_0 = d_2$, $\mu_1 = c_1 d_1 - \lambda_{11}d_2 + d_3$, $\mu_2 = \gamma_1 d_1 - \lambda_{11}d_3$, and some scalar ρ. As $\boldsymbol{G}_r\boldsymbol{l}_{n_r} = \boldsymbol{l}_{n_r}$, multiplying by $\boldsymbol{G}_r$ leaves $\boldsymbol{X}_r\mu_0 + \boldsymbol{G}_r\boldsymbol{X}_r\mu_1 + \boldsymbol{G}_r^2\boldsymbol{X}_r\mu_2$

unchanged. Thus,

$$\boldsymbol{X}_r\mu_0 + \boldsymbol{G}_r\boldsymbol{X}_r\mu_1 + \boldsymbol{G}_r^2\boldsymbol{X}_r\mu_2 = \boldsymbol{G}_r\boldsymbol{X}_r\mu_0 + \boldsymbol{G}_r^2\boldsymbol{X}_r\mu_1 + \boldsymbol{G}_r^3\boldsymbol{X}_r\mu_2. \tag{81}$$

As $\boldsymbol{I}, \boldsymbol{G}_r, \boldsymbol{G}_r^2, \boldsymbol{G}_r^3$ are linearly independent, (81) implies $\mu_0 = \mu_1 = \mu_2 = 0$, which in turn implies $d_1 = d_2 = d_3 = 0$ by the same argument as in the proof of Proposition 1. Hence, $\mathrm{E}(\boldsymbol{J}\boldsymbol{Z}_1)$ has full column rank. ■

Proposition 7 gives a condition for identification based on the variation of reference groups (i.e. an individual's direct friends) across individuals within a network. When this condition is violated, for example, all networks are complete (i.e. everybody is linked with everybody else in the network) so that $\boldsymbol{I}, \boldsymbol{G}_r, \boldsymbol{G}_r^2, \boldsymbol{G}_r^3$ are linearly dependent, identification may still be possible based on the variation across networks. For instance, Lee (2007) shows that identification can be achieved when there are variations in the network sizes. The following proposition gives a more general identification condition.[13]

Proposition 8 *Suppose* $\boldsymbol{\gamma}_1 + \lambda_{11}\boldsymbol{c}_1 \neq \boldsymbol{0}$. *Then* $\mathrm{E}(\boldsymbol{J}\boldsymbol{Z}_1)$ *of (78) has full column rank if and only if* $\boldsymbol{J}, \boldsymbol{J}\boldsymbol{G}, \boldsymbol{J}\boldsymbol{G}^2$ *are linearly independent.*

D.2 The 'triangular' system of simultaneous equations

Next, we consider the identification of

$$\begin{aligned} \boldsymbol{J}\boldsymbol{y}_1 &= \lambda_{11}\boldsymbol{J}\boldsymbol{G}\boldsymbol{y}_1 + \lambda_{21}\boldsymbol{J}\boldsymbol{G}\boldsymbol{y}_2 + \boldsymbol{J}\boldsymbol{X}\boldsymbol{c}_1 + \boldsymbol{J}\boldsymbol{G}\boldsymbol{X}\boldsymbol{\gamma}_1 + \boldsymbol{J}\boldsymbol{\epsilon}_1 & (82) \\ \boldsymbol{J}\boldsymbol{y}_2 &= \lambda_{22}\boldsymbol{J}\boldsymbol{G}\boldsymbol{y}_2 + \boldsymbol{J}\boldsymbol{X}\boldsymbol{c}_2 + \boldsymbol{J}\boldsymbol{G}\boldsymbol{X}\boldsymbol{\gamma}_2 + \boldsymbol{J}\boldsymbol{\epsilon}_2. & (83) \end{aligned}$$

Similar to the 'triangular' model without group fixed effect, a sufficient identification condition is given by the following proposition.

Proposition 9 *Suppose* $\boldsymbol{\gamma}_2 + \lambda_{22}\boldsymbol{c}_2 \neq \boldsymbol{0}$ *and* $(\lambda_{11} - \lambda_{22})(\boldsymbol{\gamma}_1 + \lambda_{11}\boldsymbol{c}_1) + \lambda_{21}(\boldsymbol{\gamma}_2 + \lambda_{11}\boldsymbol{c}_2) \neq \boldsymbol{0}$. *Then* $\mathrm{E}(\boldsymbol{J}\boldsymbol{Z}_1)$ *of model (82) has full column rank if and only if* $\boldsymbol{J}, \boldsymbol{J}\boldsymbol{G}, \boldsymbol{J}\boldsymbol{G}^2, \boldsymbol{J}\boldsymbol{G}^3$ *are linearly independent,*

D.3 The 'square' system of simultaneous equations

For the 'square' system, we consider three specifications.

[13] Proofs of Propositions 8-12 follow similar arguments as in the proofs of corresponding propositions for the models without group fixed effects.

First, we consider the 'square' model with cross-choice peer effects:

$$\boldsymbol{J}\boldsymbol{y}_1 = \lambda_{11}\boldsymbol{J}\boldsymbol{G}\boldsymbol{y}_1 + \lambda_{21}\boldsymbol{J}\boldsymbol{G}\boldsymbol{y}_2 + \boldsymbol{J}\boldsymbol{X}\boldsymbol{c}_1 + \boldsymbol{J}\boldsymbol{G}\boldsymbol{X}\boldsymbol{\gamma}_1 + \boldsymbol{J}\boldsymbol{\epsilon}_1 \tag{84}$$

$$\boldsymbol{J}\boldsymbol{y}_2 = \lambda_{22}\boldsymbol{J}\boldsymbol{G}\boldsymbol{y}_2 + \lambda_{12}\boldsymbol{J}\boldsymbol{G}\boldsymbol{y}_1 + \boldsymbol{J}\boldsymbol{X}\boldsymbol{c}_2 + \boldsymbol{J}\boldsymbol{G}\boldsymbol{X}\boldsymbol{\gamma}_2 + \boldsymbol{J}\boldsymbol{\epsilon}_2. \tag{85}$$

The following proposition gives a sufficient condition for the identification of (84). The sufficient condition for the identification of (85) can be analogously derived. Let $\boldsymbol{Z}_1 = [\boldsymbol{G}\boldsymbol{y}_1, \boldsymbol{G}\boldsymbol{y}_2, \boldsymbol{X}, \boldsymbol{G}\boldsymbol{X}]$.

Proposition 10 *If either (i) $\boldsymbol{J}, \boldsymbol{J}\boldsymbol{G}, \boldsymbol{J}\boldsymbol{G}^2, \boldsymbol{J}\boldsymbol{G}^3, \boldsymbol{J}\boldsymbol{G}^4$ are linearly independent and $\mathbf{A}_1$ given by (35) has full rank, or (ii) $\boldsymbol{J}, \boldsymbol{J}\boldsymbol{G}, \boldsymbol{J}\boldsymbol{G}^2, \boldsymbol{J}\boldsymbol{G}^3$ are linearly independent, $\boldsymbol{J}\boldsymbol{G}^4 = \rho_0\boldsymbol{J} + \rho_1\boldsymbol{J}\boldsymbol{G} + \rho_2\boldsymbol{J}\boldsymbol{G}^2 + \rho_3\boldsymbol{J}\boldsymbol{G}^3$ and $\mathbf{A}_1^*$ given by (36) has full rank, then* $\mathrm{E}(\boldsymbol{J}\boldsymbol{Z}_1)$ *of model (84) has full column rank.*

Next, we will consider the model with simultaneity effects

$$\boldsymbol{J}\boldsymbol{y}_1 = \lambda_{11}\boldsymbol{J}\boldsymbol{G}\boldsymbol{y}_1 + \beta_{12}\boldsymbol{J}\boldsymbol{y}_2 + \boldsymbol{J}\boldsymbol{X}\boldsymbol{c}_1 + \boldsymbol{J}\boldsymbol{G}\boldsymbol{X}\boldsymbol{\gamma}_1 + \boldsymbol{J}\boldsymbol{\epsilon}_1 \tag{86}$$

$$\boldsymbol{J}\boldsymbol{y}_2 = \lambda_{22}\boldsymbol{J}\boldsymbol{G}\boldsymbol{y}_2 + \beta_{21}\boldsymbol{J}\boldsymbol{y}_1 + \boldsymbol{J}\boldsymbol{X}\boldsymbol{c}_2 + \boldsymbol{J}\boldsymbol{G}\boldsymbol{X}\boldsymbol{\gamma}_2 + \boldsymbol{J}\boldsymbol{\epsilon}_2. \tag{87}$$

The following proposition gives a sufficient condition for the identification of (86). Let $\boldsymbol{Z}_1 = [\boldsymbol{G}\boldsymbol{y}_1, \boldsymbol{y}_2, \boldsymbol{X}, \boldsymbol{G}\boldsymbol{X}]$.

Proposition 11 *If either (i) $\boldsymbol{J}, \boldsymbol{J}\boldsymbol{G}, \boldsymbol{J}\boldsymbol{G}^2, \boldsymbol{J}\boldsymbol{G}^3, \boldsymbol{J}\boldsymbol{G}^4$ are linearly independent and $\mathbf{A}_2$ given by (37) has full rank, or (ii) $\boldsymbol{J}, \boldsymbol{J}\boldsymbol{G}, \boldsymbol{J}\boldsymbol{G}^2, \boldsymbol{J}\boldsymbol{G}^3$ are linearly independent, $\boldsymbol{J}\boldsymbol{G}^4 = \rho_0\boldsymbol{J} + \rho_1\boldsymbol{J}\boldsymbol{G} + \rho_2\boldsymbol{J}\boldsymbol{G}^2 + \rho_3\boldsymbol{J}\boldsymbol{G}^3$ and $\mathbf{A}_2^*$ given by (38) has full rank, then* $\mathrm{E}(\boldsymbol{J}\boldsymbol{Z}_1)$ *of model (86) has full column rank.*

Finally, for the general 'square' system with both simultaneity and cross-choice peer effects, we impose exclusion restrictions on the exogenous regressors to achieve identification. To clarify ideas, we consider the following model

$$\boldsymbol{J}\boldsymbol{y}_1 = \lambda_{11}\boldsymbol{J}\boldsymbol{G}\boldsymbol{y}_1 + \beta_{12}\boldsymbol{J}\boldsymbol{y}_2 + \lambda_{21}\boldsymbol{J}\boldsymbol{G}\boldsymbol{y}_2 + \boldsymbol{J}\boldsymbol{X}_1\boldsymbol{c}_1 + \boldsymbol{J}\boldsymbol{G}\boldsymbol{X}_1\gamma_1 + \boldsymbol{J}\boldsymbol{\epsilon}_1 \tag{88}$$

$$\boldsymbol{J}\boldsymbol{y}_2 = \lambda_{22}\boldsymbol{J}\boldsymbol{G}\boldsymbol{y}_2 + \beta_{21}\boldsymbol{J}\boldsymbol{y}_1 + \lambda_{12}\boldsymbol{J}\boldsymbol{G}\boldsymbol{y}_1 + \boldsymbol{J}\boldsymbol{X}_2\boldsymbol{c}_2 + \boldsymbol{J}\boldsymbol{G}\boldsymbol{X}_2\gamma_2 + \boldsymbol{J}\boldsymbol{\epsilon}_2, \tag{89}$$

where $\boldsymbol{X}_1$ and $\boldsymbol{X}_2$ are vectors and $\boldsymbol{X}_1 \neq \boldsymbol{X}_2$. The following proposition gives a sufficient condition for the identification of (88).

Proposition 12 *If (i)* $\boldsymbol{J}, \boldsymbol{JG}, \boldsymbol{JG}^2, \boldsymbol{JG}^3, \boldsymbol{JG}^4$ *are linearly independent and* $\mathbf{A}_3$ *given by (39) has full rank, (ii)* $\boldsymbol{J}, \boldsymbol{JG}, \boldsymbol{JG}^2, \boldsymbol{JG}^3$ *are linearly independent,* $\boldsymbol{JG}^4 = \rho_0\boldsymbol{J} + \rho_1\boldsymbol{JG} + \rho_2\boldsymbol{JG}^2 + \rho_3\boldsymbol{JG}^3$ *and* $\mathbf{A}_3^*$ *given by (40) has full rank, or (iii)* $\boldsymbol{J}, \boldsymbol{JG}, \boldsymbol{JG}^2$ *are linearly independent,* $\boldsymbol{JG}^3 = \rho_0\boldsymbol{J} + \rho_1\boldsymbol{JG} + \rho_2\boldsymbol{JG}^2$ *and* $\mathbf{A}_3^{**}$ *given by (41) has full rank, then* $\mathrm{E}(\boldsymbol{JZ}_1)$ *of (88) has full column rank.*

Table 1: Description of Data

	Label	Variable Definition	Mean	SD
Y	GPA	Average of most recent grades in English, science, math and history.	2.75	0.85
	TV	Response to the question: "During the past week, how many times did you watch television or videos, or play video games?" coded as 0 if not at all, 1 if 1 or 2 times, 2 if 3 or 4 times, 3 if 5 or more times.	2.39	0.88
X	Female	Dummy variable taking value one if the respondent is female.	0.50	0.50
	Black	Dummy variable taking value one if the respondent is black.	0.22	0.41
	Other races	Dummy variable taking value one if the respondent is neither white nor black.	0.06	0.24
	Grade	Grade of the student in the current year, coded as 1 if in grade 7, 2 if in grade 8, etc.	3.02	1.67
	Physical development	Response to the question: "How advanced is your physical development compared to other boys/girls your age?" coded as 1 if I look younger than most, 2 if I look younger than some, 3 if I look about average, 4 if I look older than some, 5 if I look older than most.	3.29	1.17
	Teacher troubles	Response to the question: "how often did you have trouble getting along with your teachers?" coded as 0 if never, 1 if just a few times, 2 if about once a week, 3 if almost everyday, 4 if everyday.	0.92	0.96
	Family size	Number of people living in the household.	2.96	2.18
	Married parents	Dummy variable taking value one if the child leaves in a family with both parents who are married.	0.51	0.50
	Parental education	Schooling level of the (biological or non-biological) parent who is living with the child, distinguishing between "never went to school", "not	3.23	1.07

		graduate from high school", "high school graduate", "graduated from college or a university", "professional training beyond a four-year college", coded as 1 to 5. If both parents are in the household the education of the father is considered. It is coded as zero if no parent lives with child or the reported level is "unknown".		
	Parent occupation manager	Parent occupation dummies. Closest description of the job of (biological or non-biological) parent that is living with the child is manager. If both parents are in the household, the occupation of the father is considered. "none" is the reference group.	0.11	0.31
	Parent occupation professional/technical	"	0.20	0.40
	Parent occupation office or sales worker	"	0.11	0.31
	Parent occupation manual	"	0.30	0.46
	Parent occupation military or security	"	0.03	0.16
	Parent occupation farm or fishery	"	0.02	0.15
	Parent occupation other	"	0.14	0.35
	Parental care	Dummy variable taking value one if both parents care very much about him/her.	0.39	0.49
	Neighborhood quality	Interviewer response to the question "How well kept is the building in which the respondent lives", coded as 4= very poorly kept (needs major repairs), 3= poorly kept (needs minor repairs), 2= fairly well kept (needs cosmetic work), 1= very well kept.	1.67	1.19
	TV watching decision	Dummy variable taking value one if the respondent's parents let him/her make own decisions about how much television to watch.	0.80	0.40

Table 2: Estimation of the Seemingly Unrelated Regression Model

	GS2SLS		GS3SLS	
	GPA	TV	GPA	TV
Gn*GPA	0.6972***		0.6955***	
	(0.1949)		(0.1949)	
Gn*TV		0.3290		0.3359
		(0.2545)		(0.2545)
Female	0.1346***	-0.1131***	0.1347***	-0.1132***
	(0.0401)	(0.0427)	(0.0401)	(0.0427)
Black	0.1130	0.0930	0.1128	0.0933
	(0.1035)	(0.1099)	(0.1035)	(0.1099)
Other races	0.0186	0.0377	0.0187	0.0385
	(0.0925)	(0.1032)	(0.0925)	(0.1032)
Grade	0.1116***	-0.0262	0.1115***	-0.0263
	(0.0266)	(0.0247)	(0.0266)	(0.0247)
Physical development	0.0342**	-0.0307*	0.0341**	-0.0306*
	(0.0165)	(0.0183)	(0.0165)	(0.0183)
Teacher troubles	-0.0691***	0.0419*	-0.0691***	0.0419*
	(0.0206)	(0.0218)	(0.0206)	(0.0218)
Family size	-0.0173	0.0189	-0.0173	0.0189
	(0.0128)	(0.0134)	(0.0128)	(0.0134)
Married parents	0.0669	-0.0543	0.0669	-0.0545
	(0.0712)	(0.0765)	(0.0712)	(0.0765)
Parental education	0.0590**	0.0005	0.0591**	0.0003
	(0.0256)	(0.0251)	(0.0256)	(0.0251)
Parental care	0.0608	0.1036	0.0609	0.1040
	(0.0636)	(0.0694)	(0.0636)	(0.0694)
Neighborhood quality	-0.0238	-0.0115	-0.0239	-0.0116
	(0.0170)	(0.0183)	(0.0170)	(0.0183)
TV watching decision	0.0044	0.1885***	0.0045	0.1885***
	(0.0494)	(0.0523)	(0.0494)	(0.0523)
Parental occupation dummies	yes		yes	
Contextual effects	yes		yes	
Network fixed effects	yes		yes	

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

Table 3: GS3SLS Estimation of the 'Triangular' Model

	Triangular Model 1		**Triangular Model 2**	
	GPA	TV	GPA	TV
Gn*GPA	0.7536***		0.6972***	0.1311
	(0.2058)		(0.1949)	(0.2104)
Gn*TV	-0.4705*	0.3290		0.3274
	(0.2516)	(0.2545)		(0.2567)
Female	0.1431***	-0.1131***	0.1346***	-0.1174***
	(0.0423)	(0.0427)	(0.0401)	(0.0432)
Black	0.0963	0.0930	0.1130	0.1037
	(0.1089)	(0.1099)	(0.1035)	(0.1112)
Other races	-0.0434	0.0377	0.0186	0.0300
	(0.1024)	(0.1032)	(0.0925)	(0.1046)
Grade	0.1198***	-0.0262	0.1116***	-0.0171
	(0.0282)	(0.0247)	(0.0266)	(0.0288)
Physical development	0.0240	-0.0307*	0.0342**	-0.0288
	(0.0181)	(0.0183)	(0.0165)	(0.0185)
Teacher troubles	-0.0664***	0.0419*	-0.0691***	0.0439**
	(0.0216)	(0.0218)	(0.0206)	(0.0221)
Family size	-0.0156	0.0189	-0.0173	0.0170
	(0.0135)	(0.0134)	(0.0128)	(0.0137)
Married parents	0.0817	-0.0543	0.0669	-0.0556
	(0.0750)	(0.0765)	(0.0712)	(0.0766)
Parental education	0.0651***	0.0005	0.0590**	-0.0071
	(0.0271)	(0.0251)	(0.0256)	(0.0277)
Parental care	0.0334	0.1036	0.0608	0.1012
	(0.0683)	(0.0694)	(0.0636)	(0.0697)
Neighborhood quality	-0.0163	-0.0115	-0.0238	-0.0096
	(0.0182)	(0.0183)	(0.0170)	(0.0186)
TV watching decision	0.0016	0.1885***	0.0044	0.1841***
	(0.0518)	(0.0523)	(0.0494)	(0.0529)
Parental occupation dummies	yes		yes	
Contextual effects	yes		yes	
Network fixed effects	yes		yes	

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

Table 4: GS3SLS Estimation of 'Square' Models 1 and 2

	Square Model 1		**Square Model 2**	
	GPA	TV	GPA	TV
Gn*GPA	0.7471***	0.1332	0.5240***	
	(0.2060)	(0.2104)	(0.1889)	
Gn*TV	-0.4695*	0.3079		0.2592
	(0.2516)	(0.2569)		(0.2594)
GPA				0.5693***
				(0.1981)
TV			0.4811***	
			(0.1694)	
Female	0.1433***	-0.1170***	0.1912***	-0.2013***
	(0.0423)	(0.0432)	(0.0445)	(0.0559)
Black	0.0958	0.1030	0.0611	0.0583
	(0.1089)	(0.1112)	(0.1055)	(0.1167)
Other races	-0.0429	0.0274	0.0301	-0.0044
	(0.1024)	(0.1046)	(0.0934)	(0.1087)
Grade	0.1194***	-0.0168	0.1108***	-0.0616**
	(0.0282)	(0.0288)	(0.0267)	(0.0293)
Physical development	0.0239	-0.0292	0.0501***	-0.0459***
	(0.0181)	(0.0185)	(0.0177)	(0.0195)
Teacher troubles	-0.0665***	0.0440**	-0.0925***	0.0876***
	(0.0216)	(0.0221)	(0.0221)	(0.0278)
Family size	-0.0155	0.0171	-0.0248*	0.0234
	(0.0135)	(0.0137)	(0.0133)	(0.0143)
Married parents	0.0817	-0.0550	0.0897	-0.0942
	(0.0750)	(0.0766)	(0.0722)	(0.0829)
Parental education	0.0655***	-0.0069	0.0657***	-0.0549
	(0.0271)	(0.0277)	(0.0257)	(0.0340)
Parental care	0.0336	0.1001	0.0230	0.0580
	(0.0683)	(0.0697)	(0.0657)	(0.0741)
Neighborhood quality	-0.0164	-0.0093	-0.0230	0.0086
	(0.0182)	(0.0186)	(0.0171)	(0.0201)
TV watching decision	0.0019	0.1839***	-0.0801	0.1723***
	(0.0518)	(0.0529)	(0.0587)	(0.0558)
Parental occupation dummies	yes		yes	
Contextual effects	yes		yes	
Network fixed effects	yes		yes	

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

Table 5: Estimation of 'Square' Model 3

	GS2SLS		GS3SLS	
	GPA	TV	GPA	TV
Gn*GPA	0.7294***	-0.1676	0.7172***	-0.4968*
	(0.2068)	(0.2993)	(0.1987)	(0.2810)
Gn*TV	-0.4817*	0.4186	-0.5565**	0.5427*
	(0.2507)	(0.2841)	(0.2492)	(0.2793)
GPA		0.3388		0.6736***
		(0.2370)		(0.2151)
TV	0.1914		0.4927***	
	(0.1630)		(0.1484)	
Female	0.1640***	-0.1629***	0.1968***	-0.2068***
	(0.0469)	(0.0551)	(0.0463)	(0.0535)
Black	0.0785	0.0614	0.0519	0.0190
	(0.1119)	(0.1201)	(0.1115)	(0.1194)
Other races	-0.0402	0.0432	-0.0477	0.0570
	(0.1036)	(0.1107)	(0.1035)	(0.1104)
Grade	0.1235***	-0.0598	0.1295***	-0.1048***
	(0.0283)	(0.0436)	(0.0281)	(0.0414)
Physical development	0.0301	-0.0387*	0.0382**	-0.0474***
	(0.0186)	(0.0205)	(0.0185)	(0.0203)
Teacher troubles	-0.0750***	0.0663***	-0.0890***	0.0884***
	(0.0233)	(0.0277)	(0.0230)	(0.0270)
Family size	-0.0192	0.0230	-0.0251*	0.0298**
	(0.0139)	(0.0151)	(0.0138)	(0.0149)
Married parents	0.0906	-0.0821	0.1100	-0.1096
	(0.0767)	(0.0828)	(0.0766)	(0.0824)
Parental education	0.0636***		0.0426*	
	(0.0271)		(0.0244)	
Parental care	0.0161	0.0882	-0.0160	0.0774
	(0.0713)	(0.0738)	(0.0710)	(0.0737)
Neighborhood quality	-0.0153	-0.0021	-0.0134	0.0037
	(0.0185)	(0.0199)	(0.0185)	(0.0198)
TV watching decision		0.1801***		0.1332***
		(0.0555)		(0.0501)
Parental occupation dummies	yes		yes	
Contextual effects	yes		yes	
Network fixed effects	yes		yes	
Overidentification test p-value	0.803	0.280		

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