Strategic Interaction in netwoks: Identification

Emerson Melo

- In this lecture we will discuss an empirical approach and some open questions in theoretical models.

Yann Bramoullé, Habiba Djebbari, Bernard Fortin, 2009, JOE

- One key challenge for the empirical literature on *peer effects* is to identify what drives the correlation between outcomes of individuals who interact together.
- Manski (1993) distinguishes between *exogenous (or contextual) effects*, i.e., the influence of exogenous peer characteristics, *endogenous effects*, i.e., the influence of peer outcomes, and *correlated effects*, i.e., individuals in the same reference group tend to behave similarly because they are alike or face a common environment.

- Manski shows that two main identification problems arise in the context of a *linear-in-means model*
- First, it is difficult to distinguish real social effects (endogenous + exogenous) from correlated effects.
- Second, even in the absence of correlated effects, simultaneity in behavior of interacting agents introduces a perfect collinearity between the expected mean outcome of the group and its mean characteristics.

- This *reflection problem* hinders the identification of the endogenous effect from the exogenous effects.
- One basic assumption that is usually made in the linear-in-means model, as well as in most peer effects models, is that individuals interact in groups.
- This means that the population is partitioned in groups, and that individuals are affected by all others in their group and by none outside of it.

- This type of interaction pattern is very particular and is not likely to represent most forms of relationship between individuals. Indeed, there is increasing recognition among economists in general of the role played by *social network* in structuring interactions among agents.
- Bramoulle et al.'s approach is inspired from the literature in spatial econometrics.
- They consider an extended version of the linear-in-means model where each individual has his own specific reference group, defined by the individuals whose mean outcome and characteristics influence his own outcome.

- Bramoulle et al. show that relaxing the assumption of group interactions generally permits to separate endogenous and exogenous effects.
- Therefore, the second negative result of Manski (1993) is not robust to reference group heterogeneity.
- Their main objective is to characterize the networks for which endogenous and exogenous effects are identifiable.
- They determine these structures both in the absence of correlated effects and when controlling for correlated effects in the form of network fixed effects.

- In both cases, they provide easy-to-check necessary and sufficient conditions for identification.
- When there are no correlated effects, tehy show that endogenous and exogenous effects are identified as soon as individuals do not interact in groups. Thus, even the slightest departure from a groupwise structure is sufficient to obtain identification.

- They consider the consumption by a secondary school student of recreational activities such as participation in artistic, sports and social organizations and clubs.
- His recreational activities are assumed to depend not only on his own characteristics (e.g., age, gender, family background), but also on his friends' mean characteristics (exogenous social effect) and their mean recreational activities (endogenous social effect).

- The latter effect may reflect conformity or simply the pleasure for the student to participate in recreational activities with friends who also participate in such activities.
- Moreover, the more (and the better) artistic, sports and social clubs are available in a school, the more likely students from this school will consume recreational activities (correlated effects).

- Using Add Health dataset, they provide estimation results on recreational activities by secondary school students when it is assumed that each individual's reference group is given by his best friends at school, as self-reported in the questionnaire.
- They show that the mean recreational activities by his friends have a positive and significant influence on a student's recreational activities.

The Basic Model

- Suppose we have a set of students i, $(i = 1, \ldots, n)$.
- Let y_i be the level of recreational activities by student i.
- Let $\mathbf{x}_i$ be a $1 \times K$ vector of characteristics of i. For simplicity, we can assume $K = 1$.
- Each student i may have a specific peers' group P_i of size n_i.
- This reference group (known by the modeler) contains all students whose recreational activities or family background may affect i's recreational activities (connections).

- Except where otherwise specified, assume that student i is excluded from his reference group, that is, $i \notin P_i$.
- **The collection of student-specific friends' groups defines a directed network between students.**
- Assume that a student's recreational activities may be affected by the mean recreational activities of his friends' group, by his family background, and by the mean family background of his friends' group.

- Formally, the structural model is given by:

$$y_i = \alpha + \beta \sum_{j \in P_i} \frac{y_i}{n_i} + \gamma x_i + \delta \sum_{j \in P_i} \frac{x_j}{n_i} + \varepsilon_i, \quad \mathbb{E}(\varepsilon_i | \mathbf{x}) = 0,$$

  where β captures the endogenous effect and δ the exogenous effect.
- It is standard to require that $|\beta| < 1$. Except for this restriction, the model does not impose any other constraints on the parameters.

- The error term ε_i reflects unobservable (to the modeler) characteristics associated with i.
- IAssume the strict exogeneity of the regressors, that is, $\mathbb{E}(\varepsilon_i|\mathbf{x}) = 0$, where $\mathbf{x}$ is an $n\times$ vector of family background.
- Thus we assume no correlated effects.
- The model is semiparametric or "distribution-free".

- The model in matrix notation can be written as:

$$\mathbf{y} = \alpha\mathbf{1} + \gamma\mathbf{x} + \delta\mathbf{G}\mathbf{x} + \varepsilon, \quad \mathbb{E}(\varepsilon|\mathbf{x}) = \mathbf{0}. \tag{1}$$

- $\mathbf{G}$ is $n \times n$ *interaction matrix* with $g_{ij} = \frac{1}{n_i}$ if i is a friend of j and 0 otherwise.

- The model in matrix notation can be written as:

$$\mathbf{y} = \alpha\mathbf{1} + \gamma\mathbf{x} + \delta\mathbf{G}\mathbf{x} + \varepsilon, \quad \mathbb{E}(\varepsilon|\mathbf{x}) = \mathbf{0}. \tag{1}$$

- $\mathbf{G}$ is $n \times n$ *interaction matrix* with $g_{ij} = \frac{1}{n_i}$ if i is a friend of j and 0 otherwise.
- The main result is to show that $\theta = (\alpha, \beta, \gamma, \delta)$ is identified given the moment restriction $\mathbb{E}(\theta|\mathbf{x}) = \mathbf{0}$ and restrictions on $\mathbf{G}$.

- We can write the restricted reduced form of model (1).
- Since $|\beta| < 1$, $\mathbf{I} - \beta\mathbf{G}$ is invertible,

$$\mathbf{y} = \alpha[\mathbf{I} - \beta\mathbf{G}]^{-1}\mathbf{1} + [\mathbf{I} - \beta\mathbf{G}]^{-1}(\gamma\mathbf{I} + \delta\mathbf{G})\mathbf{x} + [\mathbf{I} - \beta\mathbf{G}]^{-1}\varepsilon, \tag{2}$$

  where the intercept is simply $\alpha/(1-\beta)$ if the student is not isolated, and α otherwise.

Definition

We say that social effects are identified if and only if the vector θ of structural parameters can be uniquely recovered from the unrestricted reduced-form parameters in (2) (injective relationship).

- Eq (2) can be written as:

$$\mathbf{y} = \frac{\alpha}{1-\beta}\mathbf{1} + \gamma\mathbf{x} + (\gamma\beta + \delta)\sum_{k=0}^{\infty}\beta^k\mathbf{G}^{k+1}\mathbf{x} + \sum_{k=0}^{\infty}\beta^k\mathbf{G}^k\varepsilon. \tag{3}$$

- Moreover, from (3), the expected mean friends' groups' recreational activities conditional on $\mathbf{x}$ can be written as:

$$\mathbb{E}(\mathbf{G}\mathbf{y}|\mathbf{x}) = \frac{\alpha}{1-\beta}\mathbf{1} + \gamma\mathbf{G}\mathbf{x} + (\gamma\beta + \delta)\sum_{k=0}^{\infty}\beta^k\mathbf{G}^{k+2}\mathbf{x}. \tag{4}$$

Proposition 1

Suppose that $\gamma\beta + \delta \neq 0$. If the matrices $\mathbf{I}$, $\mathbf{G}$, and $\mathbf{G}^2$ are linearly independent social effects are identified. If the matrices $\mathbf{I}$, $\mathbf{G}$, and $\mathbf{G}^2$ are linearly dependent and no individual is isolated, social effects are not identified.

- The condition $\gamma\beta + \delta \neq 0$ is natural in this setting.
- As shown in (3), it means that family background of friends has some (direct and/or indirect) effect on a student's expected recreational activities.
- When is violated, endogenous and exogenous effect are zero or exactly cancel out, and social effects are absent from the reduced form. The condition is satisfied as soon as γ and δ have the same sign $\beta > 0$ and $\gamma \neq 0$.

- It can be shown that when no student is isolated, the matrices $\mathbf{I}$, $\mathbf{G}$, and $\mathbf{G}^2$ linearly dependent if and only $\mathbb{E}(\mathbf{Gy}|\mathbf{x})$ is perfectly collinear with $(\mathbf{1}, \mathbf{x}, \mathbf{Gx})$.
- This perfect collinearity means that we cannot find a valid identifying instrument for $\mathbf{Gy}$ in the structural equation (1).

- In contrast, when $\mathbb{E}(\mathbf{Gy}|\mathbf{x})$ is not perfectly collinear with the regressors, the restrictions imposed by the network structure allow the model to be identified.
- From (4), it is clear that the variables $(\mathbf{G}^2\mathbf{x}, \mathbf{G}^3\mathbf{x}, \ldots)$ can be used as identifying instruments and therefore can be used to consistently estimate the parameters.

- Suppose now that students interact through a network.
- In addition, suppose that we can find an *intransitive triad* in the network.
- Recall that this is a set of three individuals i, j;k such that i is affected by j and j is affected by k, but i is not affected by k.
- In this case $g_{ik} = 0$ while $g_{ik}^2 \geq g_{ij} g_{jk} > 0$.
- In contrast $\mathbf{G}^2 = \lambda_0 \mathbf{I} + \lambda_1 \mathbf{G}$ implies that $g_{ik}^2 = 0$.
- Therefore, the presence of an intransitive triad guarantees that $\mathbf{I}$, $\mathbf{G}$,$\mathbf{G}^2$ are linearly independent.

- This means that $\mathbf{G}^2\mathbf{x}$ is an identifying instrument for $\mathbf{Gy}$ since x_k affects y_i but only indirectly, through its effect on y_j.
- This result applies as long as there are students whose friends' friends are not all their friends.

Proposition 2

Suppose that individuals do not interact in groups. Suppose that $\gamma\beta + \delta \neq 0$. If $\mathbf{G}^2 \neq \mathbf{0}$ social effects are identified. If $\mathbf{G}^2 = \mathbf{0}$, social effects are identified when $\alpha \neq 0$, but not when $\alpha = 0$.