INSTITUT NATIONAL DE LA STATISTIQUE ET DES ETUDES ECONOMIQUES
Série des Documents de Travail du CREST
(Centre de Recherche en Economie et Statistique)

n° 2007-34

Identification of Peer Effects Using Group Size Variation

L. DAVEZIES[1]
X. D'HAULTFOEUILLE[2]
D. FOUGÈRE[3]

---

[1] DEPP, Ministère de l'Education Nationale, Paris, France and CREST-INSEE, Paris, France.

[2] ENSAE, CREST-INSEE, Paris and Université Paris I Panthéon Sorbonne, France.

[3] CNRS, CREST-INSEE, Paris, France, CEPR, London, UK and IZA, Bonn, Germany.

Identification of Peer Effects Using Group Size Variation

Laurent Davezies[a], Xavier d'Haultfoeuille[b], Denis Fougère[c]

[a] *DEPP, Ministère de l'Education Nationale, Paris, France and CREST-INSEE, Paris, France.*
[b] *ENSAE, Malakoff, France, CREST-INSEE, Paris, France and Université Paris I-Panthéon-Sorbonne, France.*
[c] *CNRS, France, CREST-INSEE, Paris, France, CEPR, London, UK and IZA, Bonn, Germany.*

November 2007

Abstract

This paper considers the semiparametric identification of endogenous and exogenous peer effects in the linear-in-means model. We show that this model is generically identified when at least three different sizes of peer groups are observed in the sample at hand. While unnecessary in general, homoskedasticity may be required in special cases to recover all parameters. Extensions to asymmetric responses to peers and binary outcomes are also considered. Once more, most parameters are semiparametrically identified under rather weak conditions. However, recovering all of them requires more stringent assumptions. Finally, we bring theoretical evidence that the model is more adapted to small groups.

JEL classification numbers: C14, C21, C25.
Keywords: social interactions; linear-in-means model; semiparametric identification.

1 Introduction

In a seminal paper, Manski (1993) showed that in a linear-in-expectations model with social interactions, endogenous and exogenous peer effects cannot be separately identified. Only a function of these two types of effects can be recovered under some strong exogeneity conditions. In the context of pupils achievement for instance, Hoxby (2000) and Ammermueller and Pischke (2006) reach identification by assuming that variations in time or between classrooms within the same school are random.[1] However, Lee (2007) has recently proposed a modified version of the social interaction model, which corresponds to a linear-in-means model, and which is shown to be identifiable without any of the previous restrictive assumptions, thanks to the group size variation.

The aim of our paper is threefold. Firstly, we re-examine the identification result proposed by Lee (2007) for the linear-in-means model. We show that the crucial assumptions here are 1) the knowledge of the group sizes, and 2) the fact that group sizes take at least three different values. On the other hand, and contrary to usual identification strategies based on reduced forms and exclusion restrictions, one does not need to observe all the members of each group. This contrasts with the usual reduced-form approach where measurement errors appear as soon as some members of the groups are missing (see Graham and Hahn, 2005). In general, neither parametric assumption nor homoskedasticity restriction on the error term are needed. Yet, we show that, in some special cases, the homoskedasticity assumption is required for recovering the structural parameters.

Secondly, we extend the analysis beyond the linear-in-means model. One important limitation of this model is the fact that the mean outcome does not depend on the allocation of individuals across groups. We consider a model with asymmetric responses to peers, and show that almost all parameters can then be identified. However, some parameters are identified only when the whole group is observed. We also consider the case of binary outcomes. Identification of discrete models with social interactions

---

[1] In the following, we will often consider the example of peer effects at school, although the model could also be applied to other topics like smoking (see e.g. Krauth, 2006), productivity in teams (see Rees et al., 2003) or retirement (Duflo and Saez, 2003).

has already been studied by, e.g., Brock and Durlauf (2001, 2007). Our model is slightly different, though, as we assume that social interactions may affect individuals through peers' latent variables rather than through their observable outcomes. This is convenient when only binary outcomes are observable, because of data limitation. To the best of our knowledge, this is the first time such a model is considered in the econometric literature. The attractive feature of our result is that it does not rely on any functional assumption concerning the errors. Yet, due to data limitation, homoskedasticity is needed to recover endogenous peer effects.

Thirdly, we discuss the theoretical background of the linear-in-means model. We show that, in the framework of a noncooperative game between members of the group, the Nash equilibrium satisfies Lee's model (2007), Manski's model (1993) or an intermediate version of these two models, depending on the amount of common knowledge of the players. Lee's model arises when their information is rich whereas Manski's model emerges when players have little information on their peers. This result shows that our model should be used when groups are small. Lastly, we also show that our model can be viewed as the stationary equilibrium of a dynamic model. Interestingly, one of the identifying assumption arises as a stability condition of this model.

The paper is organized as follows. The first section re-examines the identification of the linear-in-means model. Section two considers an extension to the case of asymmetric responses to peers. Section three is devoted to the discrete case. The fourth section discusses the theoretical foundations of the linear-in-means model. Section five concludes. Proofs are given in the appendix.

2 The linear-in-means model

Suppose that we observe R non-overlapping groups ($r = 1, ..., R$). Group r has size m_r. According to the *linear-in-means* model of Lee (2007), the outcome variable y_{ri} for individual i in group r is assumed to be a linear function of her own observable covariates, denoted x_{ri}, but also of the outcome variables and observable covariates of her peers, and of a group-specific (fixed) effect:

$$y_{ri} = x_{ri}\beta_{10} + \left(\frac{1}{m_r - 1} \sum_{j=1, j\neq i}^{m_r} y_{rj} \right) \lambda_0 + \left(\frac{1}{m_r - 1} \sum_{j=1, j\neq i}^{m_r} x_{rj} \right) \beta_{20} + \alpha_r + \varepsilon_{ri}. \tag{1}$$

Following the terminology introduced by Manski (1993), the second term in the right hand side corresponds to the endogenous peer effect, the third refers to the exogenous peer effects and α_r is a contextual (group-specific) effect. This model essentially departs from the one considered by Manski (1993) or by Graham and Hahn (2005) by replacing, on the right-hand side, the expectations relative to the whole group by the means of outcomes and covariates in the group of peers.[2] As we shall see, this small difference allows to obtain identification of the structural parameters, as soon as there is sufficient group variation.

In this section, we clarify and extend some results obtained by Lee (2007). We restrict our analysis to the case where m_r does not depend on the size of the sample. We believe that, in practice, such an assumption is virtually always satisfied. For instance, there is no reason why the mean classroom size should depend on the size of the sample. Moreover, we do not require to observe all the persons in each group. We denote n_r the number of sampled individuals in group r.

---

[2] Graham and Hahn (2005) makes the further restriction that $\beta_{20} = 0$, i.e. that there are no exogenous peer effects.

Identification is achieved by the within-group equation, that is:

$$W_{n_r}\widetilde{Y}_r = W_{n_r}\widetilde{X}_r \left(\frac{(m_r-1)\beta_{10}-\beta_{20}}{m_r-1+\lambda_0} \right) + W_{n_r}\frac{\widetilde{\varepsilon}_r}{1+\lambda_0/(m_r-1)}, \qquad (2)$$

where $\widetilde{Y}_r$ (respectively, $\widetilde{X}_r$ and $\widetilde{\varepsilon}_r$) is the vector of outcomes (respectively, of observed covariates and unobserved residuals) for individuals sampled in group r, and W_{n_r} denotes the within-group matrix of size n_r.To recover the structural parameters, we use the variation in the slope coefficient $\beta(m) = ((m-1)\beta_{10}-\beta_{20})\,/\,(m-1+\lambda_0)$. For this purpose, we make the following assumptions:

A1. For all $r = 1, ..., R$, $(\widetilde{Y}_r, \widetilde{X}_r, m_r, n_r)$ are i.i.d.[3]

A2. $\Pr(n_1 \geq 2) > 0$.

A3. $m_1, ..., m_R$ are known and should take at least three different values.

A4. For all $1 \leq i, j \leq m_1$, $E\left[x'_{1i}\ \varepsilon_{1j} \mid m_1, n_1\right] = 0$.

A5. $E\left[\widetilde{X}'_1\ W_{n_1}\ \widetilde{X}_1 \mid m_1, n_1\right]$ is almost surely nonsingular.

A6. $1 > \lambda_0 > 1 - \min\left(\mathrm{Supp}(m_1)\right)$, where $\mathrm{Supp}(m_1) = \{k \in \mathbb{N} \mid \Pr(m_1 = k) > 0\}$.

Assumption A2 simply states that the within-group approach is feasible. Assumption A3 is crucial; it imposes that group sizes are known and that these sizes vary sufficiently in the sample. Assumption A1, A4 and A5 are standard in linear panel data models, except that conditional expectations depend here both on the number of observed individuals in each group and on the group size. Conditioning by n_1 does not cause any trouble if, for instance, the observed individuals are drawn at random from the group. Finally, assumption A6 ensures that $\beta(m)$ exists for all $m \in \mathrm{Supp}(m_1)$.[4] As

---

[3] This definition is rather informal because the size of $\widetilde{Y}_r$ is random. To be rigorous, one should consider R i.i.d. infinite sequences $(\widetilde{\widetilde{Y}}_r)_{r=1...R}$ and define $\widetilde{Y}_r$ as the n_r first components of $\widetilde{\widetilde{Y}}_r$.

[4] This assumption is not minimal here, since $\lambda_0 \notin -\mathrm{Supp}(m_1 - 1)$ would be sufficient. But it is necessary in Theorems 2, 4, 5 and Lemma 1.

we will see in section 5, this assumption arises as the stability condition of a dynamic model.

Theorem 1. *Under assumptions A1-A6,* β_{10} *is identified. Moreover,*
- if $\beta_{20} \neq -\lambda_0\beta_{10}$*, then* λ_0 *and* β_{20} *are identified;*
- if $\beta_{20} = -\lambda_0\beta_{10}$*, then* λ_0 *is not identified and* β_{20} *is identified up to a constant.*

This theorem states that all parameters are generally identified provided that there is sufficient variation in the group sizes. As a notable exception, identification is lost in the absence of endogenous and exogenous peer effects (since then $\beta_{20} = -\lambda_0\beta_{10} = 0$). One can always rationalize such a model with any $\lambda_0' \neq 0$ and $\beta_{20}' = -\lambda_0'\beta_{10}$. Below, we provide a method which yields identification in this case, but it relies on a stronger assumption of homoskedasticity. In any case, one can check whether identification is lost or not, since this amounts to test whether $\beta(.)$, which is always identified, is constant.

Contrary to the reduced form approach, we do not need to know the means $(\overline{x}_r)_{1\leq r\leq R}$ in each group to identify the parameters. Thus the problem of measurement error of $\overline{x}_r$, which appears when some individuals in the group are unobserved, does not arise in our framework. Here the crucial assumption is the knowledge of the group size. If it is unknown but can be estimated, the measurement error problem comes back in a nonlinear manner. The issue of identification in this case is left for future research.[5]

Another identifying assumption is the additive nature of the group size effect. Indeed, m_r may be correlated with α_r in a general way, but we cannot add interaction terms $m_r \times x_{ri}$ to the list of regressors, since then assumption A5 would fail. Moreover, β_{10}, β_{20} and λ_0 are assumed to be

---

[5] Following Schennach (2004), the model would still be identified if two independent measures of m_r were available. The remaining issue is whether the model is identified with only one measure, as it is (under weak conditions) in a linear model (see, e.g., Lewbel, 1997).

independent of the group size. An informal test of this assumption would be to estimate $(\beta_{10}, \beta_{20}, \lambda_0)$ on strata made of groups with at least three different sizes, and to compare the estimates obtained from different strata.

If $\beta_{20} = -\lambda_0 \beta_{10}$, then λ_0 and β_{20} cannot be identified. However they can be recovered by studying variance variation under an homoskedasticity condition (assumption A7 below). More precisely, the conditional variance of the residuals should not depend on the group size. This hypothesis is rather weak since it does not imply restrictions on the form of the relationship between the residuals ε_{ri} and the covariates x_{ri}. Moreover, under A7, one needs less variability in the group sizes than previously and we can replace assumption A3 by A3'.

A3'. $m_1, ..., m_R$ are i.i.d. and known; they should take at least two different values.

A7. $\text{Var}\left(\widetilde{\varepsilon_1} \mid n_1, m_1\right) = \sigma^2 I_{n_1}$ where I_{n_1} is the identity matrix of size n_1.

Theorem 2. *Under assumptions A1-A2, A3' and A4-A7,* $(\beta_{10}, \lambda_0, \beta_{20})$ *are identified.*

The idea of using second order moments has already been used by Graham (2005) to identify peer effects. On the contrary to us, however, he can only estimate the posterior distribution of λ_0 in a Bayesian framework.

3 Asymmetric responses to peers' outcomes

One major limitation of the basic linear-in-means model is its functional form. In particular, the fact that everybody reacts similarly to peers is often considered implausible.[6] Moreover, this restriction implies that the composition of groups does not affect the mean outcome. Hence, if this

---

[6] For instance, inside the classroom, the race and gender composition variables could interact with individual race and gender. Hoxby (2000), Angrist and Lang (2004), and Cooley (2006) find that such interactions are important.

outcome corresponds to individual preferences, rearranging groups is useless for designing a public policy whose objective is to maximize a utilitarian social welfare.[7] For instance, losses due to the reallocation of students in classrooms should perfectly compensate for the gains of this reallocation. Thus, the model implies that the much debated issue of tracking versus mixing in classrooms is irrelevant in terms of efficiency.

To allow for asymmetric reactions to endogenous effects, one could consider, as Cooley (2006), a general nonlinear model. However, the identification of her model relies on the existence of a valid instrument. Instead, we propose a simple extension to the basic model (1) that does not require any exclusion restriction, because our identification strategy still works out in this framework. More precisely, we suppose that

$$y_{ri} = x_{ri}\beta_{10} + \left(\frac{1}{m_r - 1} \sum_{j=1,\, j\neq i}^{m_r} y_{rj} \right) \lambda_0(t_{ri}) + \left(\frac{1}{m_r - 1} \sum_{j=1, j\neq i}^{m_r} x_{rj} \right) \beta_{20}(t_{ri}) + \alpha_r + \varepsilon_{ri}, \tag{3}$$

where t_{ri} denotes an observable characteristic of individual i. For the sake of simplicity, we consider that t_{ri} is binary (and may take value 0 or 1).[8] The assumption that β_{10} does not depend on t_{ri} can be relaxed without loss of generality by including in x_{ri} some interaction terms between t_{ri} and the other covariates. On the other hand, an exclusion restriction is needed for identification, and we assume that x_{ri} is not reduced to t_{ri}. Let us denote x_{ri}^{-t} the components of x_{ri} different from t_{ri}.

Though simple, this model enables us to overcome the aforementioned

---

[7] However, under a social welfare function with inequality aversion, the composition of the groups does in general matter in the basic linear-in-model model.

[8] The extension to any discrete variable is straightforward. On the other hand, the case of a continuous variable is more tedious and not considered here.

drawback. Indeed, some tedious algebra gives:

$$\begin{aligned}\overline{y}_r = \; & \frac{1}{1-\left(\frac{m_r(0)\lambda_0(0)}{m_r-1+\lambda_0(0)}+\frac{m_r(1)\lambda_0(1)}{m_r-1+\lambda_0(1)}\right)} \\ & \times \left[\left(\frac{m_r(0)}{m_r\left(1+\frac{\lambda_0(0)}{m_r-1}\right)}\overline{x}_r(0)\left(\beta_{10}-\frac{\beta_{20}(0)}{m_r-1}\right)+\frac{m_r(1)}{m_r\left(1+\frac{\lambda_0(1)}{m_r-1}\right)}\overline{x}_r(1)\left(\beta_{10}-\frac{\beta_{20}(1)}{m_r-1}\right)\right)\right. \\ & \left. +\left(\frac{m_r(0)\beta_{20}(0)}{m_r-1+\lambda_0(0)}+\frac{m_r(1)\beta_{20}(1)}{m_r-1+\lambda_0(1)}\right)\overline{x}_i+\frac{m_r(0)\left(\alpha_r+\overline{\varepsilon}_r(0)\right)}{m_r\left(1+\frac{\lambda_0(0)}{m_r-1}\right)}+\frac{m_r(1)\left(\alpha_r+\overline{\varepsilon}_r(1)\right)}{m_r\left(1+\frac{\lambda_0(1)}{m_r-1}\right)}\right],\end{aligned}$$

where, for $k \in \{0,1\}$, $\overline{x}_r(k)$ denotes the mean among individuals of group r for whom $t_{ri} = k$, and where $m_r(k)$ denotes the size of this subgroup. Contrary to the basic model, it is now possible to increase the average outcome of the population, $\overline{y} = \frac{1}{m}\sum_{r=1}^{R} m_r\overline{y}_r$, by changing the $(m_r(k))_{1\leq r\leq R}$.

Let β_{10}^t (respectively, $\beta_{20}^t(t)$) denote the component of the parameter vector β_{10} (respectively, $\beta_{20}(t)$) which corresponds to the covariate denoted t. For instance, if t denotes the ethnic group of a pupil in a classroom (with, e.g., $t = 0$ for a white pupil and $t = 1$ for a black pupil), β_{10}^t denotes the impact of being black on, e.g., individual achievement, and $\beta_{20}^t(0)$ denotes the effect of the proportion of blacks in the group on the achievement of a white pupil. β_{10}^{-t} and $\beta_{20}^{-t}(t)$ correspond to the other components of parameter vectors β_{10} and $\beta_{20}(t)$. The following additional assumptions are required for identification of these parameters:

A2'. $\sup(\text{Supp}(n_1(k))) \geq 2$ for $k \in \{0,1\}$.

A5'. $E\left[\widetilde{X}_1'^{-t}(k)\; W_{n_r(k)}\; \widetilde{X}_1^{-t}(k) \mid m_1, n_1\right]$ (where $\widetilde{X}_1^{-t}(k)$ is the subvector of $\widetilde{X}_1^{-t}$ for individuals such as $t_i = k$) is almost surely nonsingular for $k \in \{0,1\}$.

A8. $\beta_{20}(k) \neq -\lambda_0(k)\beta_{10}$ for $k \in \{0,1\}$.

A9. $(\overline{x}_r, \overline{y}_r)_{1\leq r\leq R}$ are known.

Assumption A2' states that there are some subgroups in which at least two individuals are observed, so that we can use the within equation on subgroups with a positive probability. Assumption A5' supposes that variables are linearly independent within subgroups, almost surely. Assumption A8 is made for convenience. If it does not hold, one can still obtain partial identification results, as in theorem 1. Assumption A9 is restrictive, since it supposes in practice that all individuals belonging to a group are observed. However, it is not required to identify parameters β_{10}^{-t}, $\lambda_0(k)$ *and* $\beta_{20}^{-t}(k)$, $k \in \{0,1\}$.

Theorem 3. *Under assumptions A1, A2', A3, A4, A5', A6 and A8,* β_{10}^{-t}, $\lambda_0(k)$ *and* $\beta_{20}^{-t}(k)$ $(k \in \{0,1\})$ *are identified. Moreover, if assumption A9 holds,* β_{10}^{t} *and* $\beta_{20}^{t}(k)$ $(k \in \{0,1\})$ *can also be recovered.*

4 The discrete case

In this section, we investigate whether the parameters are still identified when one cannot observe directly the outcome variable but only a rough (binary) measure of it. In other terms, we observe $y_{ri} = 1\{y_{ri}^* \geq 0\}$, where y_{ri}^* satisfies equation (1). For instance, to study pupil achievement, only grade retention rather than test scores may be available. Similarly, in violence studies, only criminal (that is, sufficiently violent) acts can be observed by the econometrician. Note that these models remain essentially linear because the underlying model is linear. One could also study the case where y_{ri}^* depends on y_{rj} rather than y_{rj}^*. Such models, which have been studied by Brock and Durlauf (2001, 2007), Bayer and Timmins (2002), Tamer (2003) and Krauth (2006), are more complex since they are generally characterized by the existence of multiple equilibria.

When the outcome is a binary variable, the reduced-form equation (3) is useless for identification because $W_{n_r}\widetilde{Y}_r^*$ (where $\widetilde{Y}_r^*$ is the vector of latent outcomes for observed individuals of r) has no observational counterpart.

Instead, we rely on equation (4) below.

Lemma 1. *Suppose that* $y_{ri} = 1\{y_{ri}^* \geq 0\}$ *with* y_{ri}^* *satisfying equation (1), and that assumption A6 holds. Then the model is observationally equivalent to*

$$\begin{aligned} y_{ri} = \; & 1\left\{ x_{ri}\left(\beta_{10} - \frac{\beta_{20}}{m_r - 1}\right)\right. \\ & + \left[\overline{x_r}\frac{m_r}{m_r - 1}\left(\beta_{20} + \frac{\beta_{10}+\beta_{20}}{1-\lambda_0}\lambda_0\right) + \alpha_r\left(1 + \frac{m_r}{m_r - 1}\frac{\lambda_0}{1-\lambda_0}\right)\right] \\ & \left. + \overline{\varepsilon_r}\frac{m_r}{m_r - 1}\frac{\lambda_0}{1-\lambda_0} + \varepsilon_{ri} \geq 0 \right\}. \end{aligned} \tag{4}$$

The term between brackets corresponds to a group-specific effect. Thus we are led back to a binary model for panel data. Identification of such a model has been considered, among others, by Manski (1987), and our analysis relies on his paper. The group indices are omitted for simplifying notation. Then x_j^k denotes the k-th covariate of individual j. The following assumptions are needed for identification:

A10. $(\varepsilon_1, ..., \varepsilon_m)$ are exchangeable conditionally on $(m, x_1, ..., x_m, \alpha)$. The support of $\varepsilon_1 + \overline{\varepsilon}\frac{m}{m-1}\frac{\lambda_0}{1-\lambda_0}$ conditionally on $(m, x_1, ..., x_m, \alpha)$ is $\mathbb{R}$, almost surely.

A11. Let $z = x_2 - x_1$.[9] The support of z is not contained in any proper linear subspace of $\mathbb{R}^K$ (where K is the dimension of x_{ri}).

A12. There exists at least one component of z, for instance the k_0-th component, denoted z^{k_0}, which has everywhere a positive Lebesgue density conditional on the group size and the other components of z, i.e. conditional on the vector $(m, z^1, ..., z^{k_0-1}, z^{k_0+1}, ..., z^K)$, and whose associated slope parameter is normalized $\left(\beta_{10}^{k_0} = 1\right)$. Without loss of generality, we set $k_0 = 1$.

---

[9] Without loss of generality, we assume here that individuals 1 and 2 are observed.

The first part of assumption A10 holds for instance if $(\varepsilon_i)_{i=1,\ldots,m}$ is conditionally independent of $(x_i)_{i=1,\ldots,m}$ given m and α, and if $(\varepsilon_i)_{i=1,\ldots,m}$ are exchangeable conditionally on m, α. Hence, it is satisfied if the $(\varepsilon_i)_{i=1,\ldots,m}$ are i.i.d. and independent of $(x_1, ..., x_m, m, \alpha)$. The second part of assumption A10 is a technical condition, which is identical to the second part of assumption 1 set forth by Manski (1987). Assumption A11 ensures that z varies enough within a group. As usually in binary models, one parameter must be normalized and this is the purpose of A12. However, a small difficulty arises here, because the reduced form does not allow us to recover the sign of the structural parameters. A sufficient condition is to fix one parameter (and not only its absolute value): thus we set $\beta_{10}^1 = 1$.[10]

Theorem 4. *Suppose that assumptions A1-A3, A6 and A10-A12 hold. Then* β_{10} *is identified. Moreover,*
- if $\beta_{20} \neq \beta_{20}^1 \beta_{10}$*, then* β_{20} *is identified,*
- if $\beta_{20} = \beta_{20}^1 \beta_{10}$*,* β_{20}^1 *is not identified and the other* β_{20}^k *are identified up to* β_{20}^1*.*
On the other hand, λ_0 *is not identified.*

If fewer parameters than in model (1) are identified, Theorem 4 shows that the main attractive features of the method remain. Without any exclusion restriction and even if only two members of the groups are observed, β_{10} and β_{20} are generally identified. Similarly to the result of Theorem 1, identification of β_{20} is lost when there is no exogenous effect, because in this case $\beta_{20} = \beta_{20}^1 \beta_{10} = 0$. That λ_0 cannot be recovered is not surprising as this parameter only appears in the fixed effect and the residuals (see equation (4)). Heuristically, without any hypothesis imposed on these terms, any λ_0 can be rationalized by changing accordingly α and the $(\varepsilon_i)_{1 \leq i \leq m}$.

Thus, stronger assumptions are needed for identifying λ_0. One possibility

---

[10] Obviously, Theorem 5 also holds when $\beta_{10}^1 = -1$.

is to observe $\overline{x}$ and to restrict the dependence between the residuals and the covariates.

A3". $m_1, ..., m_R$ are i.i.d. and known; conditionally on $\overline{x}$, m_1 should take at least three values.

A9'. $\overline{x}$ is observed.

A13. $(\varepsilon_1, ..., \varepsilon_m, \alpha) \perp\!\!\!\perp (x_1, ..., x_m) \mid m, \overline{x}$.

A14. $\text{Var}(\varepsilon_1, ..., \varepsilon_m, \alpha \mid \overline{x}, m) = \begin{pmatrix} \text{Var}(\varepsilon_1 \mid \overline{x}) I_m & 0 \\ 0 & \text{Var}(\alpha \mid \overline{x}) \end{pmatrix}$.

A15. Given $(\overline{x}, m)$, the support of $\left\{ x_1(\beta_{10} - \frac{\beta_{20}}{m-1}), x_2(\beta_{10} - \frac{\beta_{20}}{m-1}) \right\}$ is $\mathbb{R}^2$.

A3" is slightly more restrictive than A3 but should hold most of the time. It is satisfied for instance when a multinomial logit (or probit) model generates m_r conditionally on $\overline{x}$. As mentioned above, assumption A9' is a restrictive condition as it imposes either to observe all individuals in the group or to consider only the covariates for which the means are known. Assumption A13 is in the same spirit than assumption A10. It restricts the dependence between α and the covariates to a dependence on the mean. Assumption A14 is the assumption of homoskedasticity in m; it is very similar to assumption A7. The difference between both assumptions stems from the identifying equation we use in both cases. In the discrete model, α remains in expression (4) and thus its variance must be specified as well as its covariance with the $(\varepsilon_i)_{1 \leq i \leq m}$.[11] Lastly, assumption A15 is a condition of large support. It especially implies that $m \geq 3$. Otherwise, indeed, the two variables belong to a line in $\mathbb{R}^2$.

Theorem 5. *Under assumptions A1-A2, A3", A6, A9', and A10-A15 and if* $\beta_{20} \neq \beta^1_{20}\beta_{10}$, λ_0 *is also identified.*

---

[11] The assumption of no covariance is not restrictive. Indeed, if the correlation between ε_i and α is not zero and independent of i, we can always reparametrize the model in order to make them uncorrelated.

5 Theoretical background

Model (1) is very close to the one considered by Manski (1993) or by Graham and Hahn (2005). Yet conclusions much differ since the basic linear-in-means model is generically identified without any exclusion restriction. Thus, one should set forth the reasons to choose this model rather than the linear-in-expectations model. An argument coming from game theory shows that this choice is in fact better justified for small groups. Indeed, we show below that identification depends on the information set of players. Equation (1) arises when players are fully informed, whereas Manski's model corresponds to an imperfect information situation. We also consider an intermediate situation where homoscedasticity is lost, and thus Theorem 2 cannot be applied.

Suppose that the utility of player i has the following form:

$$\begin{aligned}\mathcal{U}_{ri}(e_{ri},(e_{rj})_{j\neq i}) = \quad & e_{ri}\Bigg[x_{ri}\beta_{10} + \left(\frac{1}{m_r-1}\textstyle\sum_{j=1,\, j\neq i}^{m_r} e_{rj}\right)\lambda_0 \\ & + \left(\frac{1}{m_r-1}\textstyle\sum_{j=1,\, j\neq i}^{m_r} x_{rj}\right)\beta_{20} + \alpha_r + \varepsilon_{ri}\Bigg] - \tfrac{1}{2}e_{ri}^2,\end{aligned}$$

where we suppose here that the individual specific terms $(\varepsilon_{ri})_i$ are mutually independent and independent of (x_{ri}, α_r, m_r). In this game, the marginal returns of individual i depend on her peers' outcomes. This captures the fact that people are influenced by their peers' behavior. If $\lambda_0 > 0$, player i tries to conform to other members of her group, whereas she tries to stand out from them when $\lambda_0 < 0$. This model is close to the one developed by Cooley (2006) for examining pupil achievement in the classroom.

Assuming that α_r and the $(x_{ri}, \varepsilon_{ri})_{1\leq i\leq m_r}$ are observed by all players in group r, the Nash equilibrium of the game $(e_{r1}^*, ..., e_{rm_r}^*)$ satisfies equation (1). Thus, using Theorems 1 and 2, all parameters are identified provided that there is sufficient variation in group sizes.[12]

---

[12] This conclusion still holds if we do not observe the optimal effort e_{ri}^* but instead the corresponding outcome $y_{ri} = e_{ri}^* + \eta_{ri}$, where η_{ri} is an unanticipated shock, independent of other variables. A quick examination reveals that in the within-group equation (3), shocks

Now suppose that the $(\varepsilon_{rj})_{j\neq i}$ are unobserved by player i. This situation is realistic in moderately large groups where each player observes the characteristic x_{rj} of j but not her idiosyncratic shock. Then the Bayesian Nash equilibrium satisfies:[13]

$$e^*_{ri} = x_{ri}\beta_{10} + \left(\frac{1}{m_r-1}\sum_{j=1,\,j\neq i}^{m_r} E(e^*_{rj} \mid \varepsilon_{ri})\right)\lambda_0 + \left(\frac{1}{m_r-1}\sum_{j=1,\,j\neq i}^{m_r} x_{rj}\right)\beta_{20} + \alpha_r + \varepsilon_{ri}. \tag{5}$$

Taking the conditional expectation for $i \neq k$ leads to

$$E(e^*_{ri} \mid \varepsilon_{rk}) = x_{ri}\beta_{10} + \left(\frac{1}{m_r-1}\sum_{j=1,\,j\neq i}^{m_r} E(e^*_{rj})\right)\lambda_0 + \left(\frac{1}{m_r-1}\sum_{j=1,\,j\neq i}^{m_r} x_{rj}\right)\beta_{20} + \alpha_r, \tag{6}$$

because, by independence of the $(\varepsilon_{ri})_i$, $E[E(e^*_{rj} \mid \varepsilon_{ri}) \mid \varepsilon_{rk}] = E[E(e^*_{rj} \mid \varepsilon_{ri})] = E(e^*_{rj})$. Hence, taking the expectation of (5) and comparing with (6) leads to $E(e^*_{ri} \mid \varepsilon_{rk}) = E(e^*_{ri})$ for all $i \neq k$. Replacing $E(e^*_{rj} \mid \varepsilon_{ri})$ by $E(e^*_{rj})$ in (5), we get $e^*_{ri} = E(e^*_{ri}) + \varepsilon_{ri}$. Finally, substituting $e^*_{rj} - \varepsilon^*_{rj}$ to $E(e^*_{rj})$ in (5), we obtain:

$$\begin{aligned} e^*_{ri} = \quad & x_{ri}\beta_{10} + \left(\frac{1}{m_r-1}\sum\nolimits_{j=1,\,j\neq i}^{m_r} e^*_{rj}\right)\lambda_0 + \left(\frac{1}{m_r-1}\sum\nolimits_{j=1,\,j\neq i}^{m_r} x_{rj}\right)\beta_{20} \\ & + \alpha_r - \frac{m_r}{m_r-1}\overline{\varepsilon_r}\lambda_0 + \varepsilon_{ri}\left(1 + \frac{\lambda_0}{m_r-1}\right), \end{aligned}$$

which is equivalent to

$$e^*_{ri} = x_{ri}\beta_{10} + \left(\frac{1}{m_r-1}\sum_{j=1,\,j\neq i}^{m_r} e^*_{rj}\right)\lambda_0 + \left(\frac{1}{m_r-1}\sum_{j=1,\,j\neq i}^{m_r} x_{rj}\right)\beta_{20} + \alpha'_r + \varepsilon'_{ri},$$

where $\alpha'_r = \alpha_r - m_r/(m_r-1)\overline{\varepsilon_r}\lambda_0$ and $\varepsilon'_{ri} = \varepsilon_{ri}\left(1 + \frac{\lambda_0}{m_r-1}\right)$. This equation is very similar to (1) and the parameters can be identified using the exogeneity of covariates. However, the errors become heteroskedastic. More

---

are not homoskedastic anymore, but the proof of Theorem 2 can be adapted provided that at least three different group sizes are available.

[13] All expectations are in fact taken conditionally on $(x_{ri})_{1\leq i\leq m_r}$ and α_r. We omit them for the sake of simplicity.

precisely, the variance of residuals in the within-group equation (2) is no more dependent on m_r, so that λ_0 cannot be recovered by using this device, as in Theorem 2. Identification of β_{20} and λ_0 is lost when $\beta_{20} = -\lambda_0\beta_{10}$.

Lastly, let us suppose that the $(x_{rj})_{j\neq i}$ are also unobserved by i but that α_r is still observed. If groups are large, player i may not know j and thus does not observe neither ε_{rj} nor x_{rj}. On the other hand, she may know the general features of the group, represented by α_r and $E(x_{r1} \mid \alpha_r)$. Then, proceeding as previously, we can easily show that

$$e^*_{ri} = x_{ri}\beta_{10} + E(x_{r1} \mid \alpha_r)\beta_{20} + E(e^*_{ri} \mid \alpha_r)\lambda_0 + \alpha_r + \varepsilon_{ri}.$$

Thus we are led back to Manski's model (1993), which is not identifiable.

Another structural interpretation of the model is to consider equation (1) as the stationary state of a dynamic model with interactions. Let y^t_{ri} denote the outcome of individual i in group r at date t, and suppose that the variation in this outcome at date t is generated by the following first-order differential equation:

$$\frac{dy^t_{ri}}{dt} = y^t_{ri}\mu + \left(\frac{1}{m_r-1}\sum_{j=1,\, j\neq i}^{m_r} y^t_{rj}\right)\lambda_0 + x_{ri}\beta_{10} + \left(\frac{1}{m_r-1}\sum_{j=1,\, j\neq i}^{m_r} x_{rj}\right)\beta_{20} + \alpha_r + \varepsilon_{ri}. \tag{7}$$

In other terms, the variation in the i-th individual's outcome at date t depends on her current outcome and on her time-constant observable covariates, but also on the mean of her peers' outcomes at date t and on the mean of their time-constant observable covariates. In the example of a classroom, $\lambda_0 > 0$ implies that a pupil will get better test scores when the mean score of her classmates is higher.

Proposition 1. *Equation (7) admits a stable solution for all $r = 1...R$ if and only if*

$$\mu \leq 0 \text{ and } \lambda_0 \in \left]\mu\left\{\min_r(m_r) - 1\right\}, -\mu\right[.$$

Provided that $\mu = -1$, this solution is given by equation (1).

Setting $\mu = -1$ is natural in view of identification. Indeed, in the steady state, $dy_{ri}^{t}/dt = 0$ so that equation (7) can be normalized without loss of generality. Proposition 1 states that if the dynamic model is true, then λ_0 lies in a known interval. Interestingly, this restriction ($\lambda_0 > 1 - \min_r(m_r)$) is used both in Lee's paper (assumption 3) and in ours (assumption A6) to identify the model.

6 Conclusion

This paper considers identification of linear social interaction models using group size variation. Provided that the size of the group is known and varies sufficiently, endogenous and exogenous effects can be identified without any exclusion restriction. Moreover, the method does not require to observe all the members of each group. The result is extended to asymmetric models and binary outcomes. We also show that, from a game theoretic point of view, our analysis is more relevant for small groups such as classrooms. When groups are large, Manski's model (1993) should be more appropriate in terms of players' information. The result is also in line with the weak identification result obtained by Lee (2007) in the case of large groups.

Our paper has two main limitations. First, the size of the group is assumed to be known. However, as emphasized by Manski (2000), it is often difficult to define groups on an a priori background. This criticism is common to all models of social interactions, but may be especially problematic here. Indeed, ignoring the boundaries of the group leads (among other difficulties) to measurement errors on the group size, which could prevent identification. Second, we do not consider a fully nonparametric regression. The issue of whether group size variation has an identifying power in this general case remains to be settled.

Acknowledgements

We would like to thank Stéphane Grégoir, Steve Machin, Amine Ouazad, as well as the participants of the econometric seminar of the CREST and of the CEPR summer school in Padova for their useful comments. The usual disclaimer applies.

Appendix

Theorem 1

First, under assumption A4, $E(\widetilde{X}_1' W_{n_1} \widetilde{\varepsilon}_1 \mid n_1, m_1) = 0$ and thus, by assumption A5, $\beta(m)$ is identified for all $m \in \text{Supp}(m_1)$. We now prove that the knowledge of $m \mapsto \beta(m)$ allows in general to recover the structural parameters.

Let $(m_1^*, m_2^*) \in \text{Supp}(m_1)^2$, then

$$\frac{(m_1^* - 1)\beta_{10} - \beta_{20}}{m_1^* - 1 + \lambda_0} = \frac{(m_2^* - 1)\beta_{10} - \beta_{20}}{m_2^* - 1 + \lambda_0}$$

is equivalent to

$$-\lambda_0 \beta_{10} \left(\frac{1}{m_1^* - 1} - \frac{1}{m_2^* - 1} \right) = \beta_{20} \left(\frac{1}{m_1^* - 1} - \frac{1}{m_2^* - 1} \right).$$

Hence, if $\beta_{20} = -\lambda_0 \beta_{10}$, $\beta(.)$ is constant, and if not, $\beta(.)$ is a one-to-one mapping. In the first case, $\beta(m) = \beta_{10}$ for all m. Thus β_{10} is identified, but λ_0 cannot be recovered by $\beta(.)$. Because $\beta_{20} = -\lambda_0 \beta_{10}$, β_{20} is identified up to a constant.

Now suppose that $\beta_{20} \neq -\lambda_0 \beta_{10}$ and let (m_0^*, m_1^*, m_2^*) be three different values in $\text{Supp}(m_1)$. We will prove that the knowledge of $\beta(m_0^*), \beta(m_1^*)$ and $\beta(m_2^*)$ permits to recover $(\beta_{10}, \lambda_0, \beta_{20})$. This amounts to show that the system

$$\begin{cases} \beta(m_0^*)\lambda_0 - (m_0^*-1)\beta_{10} + \beta_{20} & = -\beta(m_0^*)(m_0^*-1) \\ \beta(m_1^*)\lambda_0 - (m_1^*-1)\beta_{10} + \beta_{20} & = -\beta(m_1^*)(m_1^*-1) \\ \beta(m_2^*)\lambda_0 - (m_2^*-1)\beta_{10} + \beta_{20} & = -\beta(m_2^*)(m_2^*-1) \end{cases}$$

has a unique solution. Using the matrix form, we can rewrite the system as $A\zeta_0 = B$ where $\zeta_0 = (\lambda_0, \beta_{10}, \beta_{20})'$. If $\det(A) \neq 0$, ζ_0 is identified. Suppose that $\det(A) = 0$. Then $com(A)'B = 0$ where $com(A)$ denotes the comatrix of A. By using the first line of this equation and the expression of $det(A)$, we get

$$\begin{cases} (m_2^*-m_1^*)\beta(m_0^*) + (m_0^*-m_2^*)\beta(m_1^*) + (m_1^*-m_0^*)\beta(m_2^*) = 0 \\ (m_0^*-1)(m_2^*-m_1^*)\beta(m_0^*) + (m_1^*-1)(m_0^*-m_2^*)\beta(m_1^*) + (m_2^*-1)(m_1^*-m_0^*)\beta(m_2^*) = 0. \end{cases}$$

Hence,

$$\begin{cases} (m_2^*-m_1^*)\beta(m_0^*) = -(m_0^*-m_2^*)\beta(m_1^*) - (m_1^*-m_0^*)\beta(m_2^*) \\ m_0^*(m_2^*-m_1^*)\beta(m_0^*) + m_1^*(m_0^*-m_2^*)\beta(m_1^*) + m_2^*(m_1^*-m_0^*)\beta(m_2^*) = 0. \end{cases}$$

Thus,

$$\begin{cases} (m_2^*-m_1^*)\beta(m_0^*) + (m_0^*-m_2^*)\beta(m_1^*) + (m_1^*-m_0^*)\beta(m_2^*) = 0 \\ \beta(m_1^*)(m_0^*-m_2^*)(m_1^*-m_2^*) + \beta(m_0^*)(m_2^*-m_1^*)(m_0^*-m_2^*) = 0. \end{cases}$$

Because $m_1^* \neq m_2^*$ and $m_0^* \neq m_2^*$, this implies that $\beta(m_1^*) = \beta(m_0^*)$, which is in contradiction with the fact that $\beta(.)$ is a one-to-one mapping. Thus $\det(A) \neq 0$ and ζ_0 is identified.

Theorem 2

Because $m \mapsto \beta(m)$ is identified, $\text{Var}\left(\dfrac{W_{n_1}\widetilde{\varepsilon}_1}{1+\dfrac{\lambda_0}{m_1-1}} \mid n_1, m_1\right)$ is known. Thus, under assumption A7,

$$\text{Var}\left(\frac{W_{n_1}\widetilde{\varepsilon}_1}{1+\frac{\lambda_0}{m-1}} \mid n_1, m_1\right) = \frac{\sigma^2}{\left(1+\frac{\lambda_0}{m_1-1}\right)^2} W_{n_1}.$$

Hence, for $m_1^* \neq m_2^*$,

$$C \equiv \frac{\left(1+\frac{\lambda_0}{m_1^*-1}\right)^2}{\left(1+\frac{\lambda_0}{m_2^*-1}\right)^2}$$

is identified. Under assumption A6, $\left(1+\frac{\lambda_0}{m-1}\right) > 0$ for all $m \in \text{Supp}(m_1)$. Thus

$$\left(\frac{\sqrt{C}}{m_1^*-1} - \frac{1}{m_2^*-1}\right)\lambda_0 = 1-\sqrt{C}.$$

It is clear that $\left(\frac{\sqrt{C}}{m_1^*-1} - \frac{1}{m_2^*-1}\right) \neq 0$. Otherwise $C=1$ and then $m_1^* = m_2^*$, which contradicts the assumption. Thus λ_0 is identified.

Then, because $m \mapsto \beta(m)$ is identified, $\beta_{10} - \dfrac{\beta_{20}}{m-1}$ is known for all $m \in \text{Supp}(m_1)$. Taking two different values for m allows to recover β_{20}, and then β_{10}.

Theorem 3

Let i_1 and i_2 be two individuals in group r such as $t_{ri_1} = t_{ri_2} = k$. Then

$$y_{ri_1} - y_{ri_2} = (x_{ri_1} - x_{ri_2})\left(\frac{(m_r-1)\beta_{10} - \beta_{20}(k)}{m_r - 1 + \lambda_0(k)}\right) + \frac{\varepsilon_{ri_1} - \varepsilon_{ri_2}}{1+\lambda_0(k)/(m_r-1)}.$$

Thus, by application of Theorem 1 on both subgroups, $\beta_{10}^{-t}, \beta_{20}^{-t}(k)$ and $\lambda_0(k)$ are identified. To recover β_{10}^t and $\beta_{20}^t(k)$, we start from

$$\begin{aligned} y_{ri}\left(m_r - 1 + \lambda_0(t_{ri})\right) &- x_{ri}^{-t}\left((m_r-1)\beta_{10}^{-t} - \beta_{20}^{-t}(t_{ri})\right) - m_r\overline{y}_r\lambda_0(t_{ri}) + m_r\overline{x}_r^{-t}\beta_{20}^{-t}(t_{ri}) \\ &= t_{ri}\left((m_r-1)\beta_{10}^t - \beta_{20}^t(t_{ri})\right) + m_r\overline{t}_r\beta_{20}^t(t_{ri}) + (m_r-1)\alpha_r + (m_r-1)\varepsilon_{ri}. \end{aligned}$$

Let $\widehat{y}_{ri}$ denote the left hand side of this equation; it is identified thanks to the previous result and to assumption A9. Taking i_1 and i_3 such as $t_{ri_1} = 1$, $t_{ri_3} = 0$, we get

$$\widehat{y}_{ri_1} - \widehat{y}_{ri_3} = (m_r-1)\beta_{10}^t - \beta_{20}^t(1) + m_r\overline{t}_r\left[\beta_{20}^t(1) - \beta_{20}^t(0)\right] + (m_r-1)(\varepsilon_{ri_1} - \varepsilon_{ri_3}).$$

This regression (over groups of the same size but with different $\overline{t}_r$) enables us to recover the constant $(m_r-1)\beta_{10}^t - \beta_{20}^t(1)$ and $\beta_{20}^t(1) - \beta_{20}^t(0)$. Then making m_r vary allows us to identify β_{10}^t and $\beta_{20}^t(1)$, and thus $\beta_{20}^t(0)$.

Lemma 1

Applying the between-group operator to (1) gives

$$\overline{y_r^*} = \overline{x_r}\left(\frac{\beta_{10}+\beta_{20}}{1-\lambda_0}\right) + \frac{\alpha_r}{1-\lambda_0} + \frac{\overline{\varepsilon_r}}{1-\lambda_0},$$

since $1/(1-\lambda_0)$ exists, according to assumption A6. Consequently, replacing $\overline{y_r^*}$ in equation (1), we obtain

$$\begin{aligned} y_{ri}^*\left(1+\frac{\lambda_0}{m_r-1}\right) &= x_{ri}\left(\beta_{10}-\frac{\beta_{20}}{m_r-1}\right) + \overline{x_r}\frac{m_r}{m_r-1}\left(\beta_{20}+\frac{\beta_{10}+\beta_{20}}{1-\lambda_0}\lambda_0\right) \\ &\quad +\alpha_r\left(1+\frac{m_r}{m_r-1}\frac{\lambda_0}{1-\lambda_0}\right) + \overline{\varepsilon_r}\frac{m_r}{m_r-1}\frac{\lambda_0}{1-\lambda_0} + \varepsilon_{ri}. \end{aligned}$$

Note that this equation is equivalent to equation (1). Now, under assumption A6, $1+\lambda_0/(m_r-1) > 0$ for all r, so that $y_{ri}^* \geq 0$ if and only if $y_{ri}^*\left(1+\frac{\lambda_0}{m_r-1}\right) \geq 0$. Thus, under assumption A6, $y_{ri} = 1\{y_{ri}^* \geq 0\}$, where y_{ri}^* satisfies equation (1), is observationally equivalent to y_{ri} satisfying equation (4).

Theorem 4

Assumption A10 implies that the conditional distribution of $\overline{\varepsilon}\frac{m}{m-1}\frac{\lambda_0}{1-\lambda_0}+\varepsilon_i$ is the same for each i. Thus assumption 1 in Manski (1987) is satisfied and, using A11 and A12, we can apply directly his result to recover $\frac{(m-1)\beta_{10}-\beta_{20}}{|m-1-\beta_{20}^1|}$. The first term of the vector, $\frac{(m-1)\beta_{10}^1-\beta_{20}^1}{|m-1-\beta_{20}^1|}$, is also identified. By assumption A12,

$$\widetilde{\beta}(m) \equiv \frac{(m-1)\beta_{10}-\beta_{20}}{m-1-\beta_{20}^1} = \left(\frac{(m-1)\beta_{10}-\beta_{20}}{|m-1-\beta_{20}^1|}\right) / \left(\frac{(m-1)\beta_{10}^1-\beta_{20}^1}{|m-1-\beta_{20}^1|}\right),$$

so that $\widetilde{\beta}(m)$ is identified as the ratio of two known terms. The rest of the proof of identification of (β_{10},β_{20}) follows the same line than the one of Theorem 1, λ_0 being replaced by $\left(-\beta_{20}^1\right)$.

However, λ_0 cannot be identified. Indeed, let $\lambda_0' \neq \lambda_0$ and define

$$\varepsilon_i' = \varepsilon_i + \overline{\varepsilon}\frac{m(\lambda_0 - \lambda_0')}{(m-1+\lambda_0')(1-\lambda_0)}.$$

Finally let

$$\alpha' = \frac{m\overline{x}(\beta_{10}+\beta_{20})(\lambda_0 - \lambda_0') + \alpha(m-1+\lambda_0)(1-\lambda_0')}{(m-1+\lambda_0')(1-\lambda_0)}.$$

Then $(\lambda_0', \alpha', \varepsilon_1', ..., \varepsilon_m')$ are observationally equivalent to the initial model. Indeed, we can check that they lead to equation (4) as well. Moreover, conditioning on $(m, x_1, ..., x_m, \alpha')$ is equivalent to conditioning on $(m, x_1, ..., x_m, \alpha)$, and conditional exchangeability of $(\varepsilon_1, ..., \varepsilon_m)$ implies conditional exchangeability of the $(\varepsilon_1', ..., \varepsilon_m')$. Furthermore, letting F_u denote the c.d.f. of any random variable u,

$$F_{\varepsilon_1' + \overline{\varepsilon'}\frac{m}{m-1}\frac{\lambda_0'}{1-\lambda_0'}|m=m^*, x_1=x_1^*, ..., x_m=x_m^*, \alpha'=\alpha'^*} = F_{\varepsilon_1 + \overline{\varepsilon}\frac{m}{m-1}\frac{\lambda_0}{1-\lambda_0}|m=m^*, x_1=x_1^*, ..., x_m=x_m^*, \alpha=\alpha^*},$$

where

$$\alpha^* = \frac{(m-1+\lambda_0')(1-\lambda_0)\alpha'^* - m\overline{x}(\beta_{10}+\beta_{20})(\lambda_0-\lambda_0')}{(m-1+\lambda_0)(1-\lambda_0')}.$$

Thus the second part of assumption A10 also holds with $(\lambda_0', \alpha', \varepsilon_1', ..., \varepsilon_m')$. This shows that λ_0 is not identified.

Theorem 5

Let $\theta_0 = \dfrac{\lambda_0}{1-\lambda_0}$ and

$$v_i = \left[\overline{x}\tfrac{m}{m-1}\left[\beta_{20} + \theta_0(\beta_{10}+\beta_{20})\right] + \alpha\left(1 + \tfrac{m}{m-1}\theta_0\right)\right] + \overline{\varepsilon}\tfrac{m}{m-1}\theta_0 + \varepsilon_i.$$

Note that $F_{v_1,...,v_m|x_1,...,x_m,m} = F_{v_1,...,v_m|\overline{x},m}$. Indeed

$$F_{v_1,...,v_m|x_1,...,x_m,m}(v_1^*,...,v_m^*|x_1^*,...,x_m^*,m^*)$$

$$= \int F_{v_1,...,v_m|x_1,...,x_m,m,\alpha}(v_1^*,...,v_m^*|x_1^*,...,x_m^*,m^*,\alpha^*)dF_{\alpha|x_1,...,x_m,m}(\alpha^*|x_1^*,...,x_m^*,m^*)$$

$$= \int F_{v_1,...,v_m|\overline{x},\alpha,m}(v_1^*,...,v_m^*|\overline{x^*},\alpha^*,m^*)dF_{\alpha|\overline{x},m}(\alpha^*|\overline{x^*},m)$$

$$= F_{v_1,...,v_m|\overline{x},m}(v_1^*,...,v_m^*|\overline{x^*},m^*),$$

where the third line stems from assumption A13 and the fact that, given $x_1,...,x_m,m,\alpha$, $(v_1,...,v_m)$ is a deterministic function of $(\varepsilon_1,...,\varepsilon_m)$. Now

$$\Pr(y_1 = 0, y_2 = 0 \mid x_1 = x_1^*, x_2 = x_2^*, \overline{x} = x, m = m^*)$$

$$= \Pr\left\{v_1 \leq -x_1^*\left(\beta_{10} - \tfrac{\beta_{20}}{m-1}\right),\ v_2 \leq -x_2^*\left(\beta_{10} - \tfrac{\beta_{20}}{m-1}\right) \mid x_1 = x_1^*, x_2 = x_2^*, \overline{x} = x, m = m^*\right\}$$

$$= F_{v_1,v_2|\overline{x},m}\left(-x_1^*\left(\beta_{10} - \tfrac{\beta_{20}}{m^*-1}\right), -x_2^*\left(\beta_{10} - \tfrac{\beta_{20}}{m^*-1}\right)|x,m^*\right).$$

Because, by Theorem 4, (β_{10},β_{20}) is identified, $x_1^*\left(\beta_{10} - \tfrac{\beta_{20}}{m^*-1}\right)$ and $x_2^*\left(\beta_{10} - \tfrac{\beta_{20}}{m^*-1}\right)$ are known. Moreover, $\overline{x}$ is observed so that the first term is identified on the whole support of (x_1, x_2). Thus, by assumption A15, making (x_1, x_2) vary allows us to recover the whole conditional distribution of (v_1, v_2) given $\overline{x}$ and m. Thus, using assumption A14,

$$\text{Cov}(v_1, v_1 - v_2 \mid \overline{x}, m) = \text{Cov}\left(\overline{\varepsilon}\frac{m}{m-1}\theta_0 + \varepsilon_1, \varepsilon_1 - \varepsilon_2 \mid \overline{x}, m\right) = \text{Var}(\varepsilon_1 \mid \overline{x}),$$

so that the r.h.s. term is identified. Let us note that assumption A14 can be tested thanks to the data, since it implies that $\text{Cov}(v_1, v_1 - v_2 \mid \overline{x}, m)$

does not depend on m. Moreover, a little algebra shows that

$$\begin{aligned}(m-1)^2\text{Cov}(v_1, v_2|\overline{x}, m) \quad &= \quad m^2\left[(1+\theta_0)^2\text{Var}(\alpha|\overline{x})\right] + m\Bigg[-2(1+\theta_0)\text{Var}(\alpha|\overline{x}) \\ &\qquad +\theta_0(2+\theta_0)\text{Var}(\varepsilon_1|\overline{x})\Bigg] + \left[\text{Var}(\alpha|\overline{x}) - 2\theta_0\text{Var}(\varepsilon_1|\overline{x})\right].\end{aligned}$$

Conditionally on $\overline{x}$, this is a regression of the (known) l.h.s. term on $(m^2, m, 1)$. By $A3''$, there exists a set A of positive probability such that m can take three different values with positive probability conditionally on $\overline{x} = x^*$, for all $x^* \in A$. Thus, the coefficients (a, b, c) of this regression can be recovered.[14] We will show that the knowledge of these coefficients implies that θ_0 is identified. The conclusion will follow because θ_0 is one-to-one with λ_0.

First, set $\phi_0 = 1 + \theta_0$ and $\rho_0 =$Var$(\alpha|\overline{x})/$Var$(\varepsilon|\overline{x})$. Moreover we set $a' = a/\text{Var}(\varepsilon|\overline{x})$, $b' = b/\text{Var}(\varepsilon|\overline{x}) + 1$ and $c' = c/\text{Var}(\varepsilon|\overline{x}) - 2$. Then a', b' and c' are identified, and

$$\left\{\begin{array}{rcl} \phi_0^2\rho_0 & = & a' \\ -2\phi_0\rho_0 + \phi_0^2 & = & b' \\ \rho_0 - 2\phi_0 & = & c' \end{array}\right.$$

Replacing ρ_0 by $c' + 2\phi_0$ in the first and second equation leads to

$$\left\{\begin{array}{rcl} \phi_0^3 + c'/2\phi_0^2 - a'/2 & = & 0 \\ \phi_0^2 + 2c'/3\phi_0 + b'/3 & = & 0 \\ \rho_0 - 2\phi_0 & = & c' \end{array}\right. \tag{8}$$

This system admits at most two solutions in (ρ, ϕ). Suppose that there are two different solutions, and let (ρ_1, ϕ_1) denote the second one. Then we can write the polynomial of the first equation as a product in which one factor is the polynomial of the second equation. Hence, there exists x such as, for all $\phi \in \mathbb{R}$,

$$\phi^3 + c'/2\phi^2 - a'/2 = (\phi^2 + 2c'/3\phi^2 + b'/3)(\phi + x).$$

---

[14] These coefficients depend on $\overline{x}$ but for the sake of simplicity, we keep this dependency implicit in the sequel.

Thus

$$\left\{\begin{array}{rcl} x &=& -c'/6 \\ 2c'x &=& -b' \\ 2b'x &=& -3a' \end{array}\right.$$

Hence $c'^2 = 3b'$. Replacing b' and c' by their expression gives

$$3(-2\phi\rho + \phi^2) = (\rho - 2\phi)^2,$$

which must hold for (ρ_0, ϕ_0) and (ρ_1, ϕ_1). But this statement is equivalent to $\phi + \rho = 0$. Replacing ρ by $-\phi$ in c' gives $\phi_0 = \phi_1 = -c'/3$ and thus also $\rho_0 = \rho_1$. This contradicts $(\rho_0, \phi_0) \neq (\rho_1, \phi_1)$. Thus ϕ_0 is identified by (8) and the conclusion follows.

Proposition 1

Let $k_{ri} = x_{ri}\beta_{10} + \left(\frac{1}{m_r-1}\sum_{j=1,\, j\neq i}^{m_r} x_{rj}\right)\beta_{20} + \alpha_r + \varepsilon_{ri}$, $Y_r^t = (y_{r1}^t, ..., y_{rm_r}^t)'$ and $K_r = (k_{r1}, ..., k_{rm_r})'$. Equation (2) can be rewritten in matrix terms:

$$\frac{dY_r^t}{dt} = \left[\left(\mu - \frac{\lambda_0}{m_r - 1}\right) I_{m_r} + \frac{\lambda_0}{m_r - 1} J_{m_r}\right] Y_r^t + K_r,$$

where J_{m_r} denotes the matrix of ones of size m_r. The eigenvalues of $A_r = (\mu - \frac{\lambda_0}{m_r-1}) I_{m_r} + \frac{\lambda_0}{m_r-1} J_{m_r}$ are $\left(\mu - \frac{\lambda_0}{m_r-1}\right)$ and $(\mu + \lambda_0)$. Thus the system is stable when $\max(\mu - \frac{\lambda_0}{m_r-1}, \mu + \lambda_0) \leq 0$. Because $\max(\mu - \frac{\lambda_0}{m_r-1}, \mu + \lambda_0) \geq \mu$, this implies that $\mu \leq 0$. Moreover, when $\lambda_0 \geq 0$, the condition is equivalent to $\lambda_0 \leq -\mu$ and when $\lambda_0 < 0$, this amounts to $\lambda_0 \geq \mu(m_r - 1)$. Thus, the system is stable for all $r = 1, ..., R$, if and only if

$$\mu \leq 0 \ \text{ and } \ \lambda_0 \in [\mu\{\min_r(m_r) - 1\}, -\mu].$$

If these conditions are fulfilled, then, $\lim_{t\to+\infty} A_r Y_r^t = -K_r$. Equivalently, if y_{ri} denotes the i-th component of the stationary state of Y_r^t,

$$-\mu y_{ri} = \left(\frac{1}{m_r - 1}\sum_{j=1,\, j\neq i}^{m_r} y_{rj}\right)\lambda_0 + x_{ri}\beta_{10} + \left(\frac{1}{m_r - 1}\sum_{j=1,\, j\neq i}^{m_r} x_{rj}\right)\beta_{20} + \alpha_r + \varepsilon_{ri}.$$

Thus, when $\mu = -1$, we retrieve equation (1).

References

Ammermueller A. and J. S. Pischke, 2006, Peer Effects in European Primary Schools: Evidence from PIRLS. IZA Discussion Paper No. 2077.

Angrist J. D. and K. Lang, 2004, Does School Integration Generate Peer Effects? Evidence from Boston's Metco Program. The American Economic Review 94, 1613-1634.

Bayer, P. and C. Timmins, , 2002, Identifying Social Interactions in Endogenous Sorting Models. Working Paper, Department of Economics, Yale University.

Brock W. A. and S. M. Durlauf, 2001, Discrete Choice with Social Interactions. Review of Economic Studies 68, 235-260.

Brock W. A. and S. M. Durlauf, 2007, Identification of Binary Choice Models with Social Interactions. Journal of Econometrics 140, 52-75.

Cooley J., 2006, Desegregation and the Achievement Gap: Do Diverse Peers Help? Working Paper, Duke University.

Duflo E. and E. Saez, 2003, The Role of Information and Social Interactions in Retirement Plan Decisions: Evidence from a Randomized Experiment. Quarterly Journal of Economics 118, 815-842.

Graham B. S., 2005, Identifying Social Interactions through Excess Variance Contrasts. Working Paper, Harvard University.

Graham B. S. and J. Hahn, 2005, Identification and Estimation of the Linear-in-Means Model of Social Interactions. Economics Letters 88, 1-6.

Hoxby C., 2000, Peer Effects in the Classroom: Learning from Gender and Race Variation. NBER Working Paper No 7867.

Krauth B. V., 2006, Simulation-Based Estimation of Peer Effects. Journal of Econometrics 133, 243-271.

Lee L. F., 2007, Identification and Estimation of Econometric Models with Group Interactions, Contextual Factors and Fixed Effects. Journal of Econometrics 140, 333-374.

Lewbel A., 1997, Constructing Instruments for Regressions With Measurement Error When no Additional Data are Available, with an Application to Patents and R & D. Econometrica 65, 1201-1213.

Lewbel A., 2000, Semiparametric Qualitative Response Model Estimation With Unknown Heteroscedasticity and Instrumental Variables. Journal of Econometrics 97, 145-177.

Manski C., 1987, Semiparametric Analysis of Random Effects Linear Models from Binary Panel Data. Econometrica 55, 357-362.

Manski C., 1993, Identification of Endogenous Social Effects: the Reflection Problem. Review of Economic Studies 60, 531-542.

Manski C., 2000, Economic Analysis of Social Interactions. Journal of Economic Perspectives 14, 115-136.

Rees D. I., J. S. Zax and J. Herries, 2003, Interdependence in Worker Productivity. Journal of Applied Econometrics 18, 585-604.

Schennach S., 2004, Estimation of Nonlinear Models with Measurement Error. Econometrica 72, 33-75.

Tamer E., 2003, Incomplete Simultaneous Discrete Response Models with Multiple Equilibria. Review of Economic Studies 70, 147–165.