Identification of Peer Effects Using Group Size Variation*

Laurent Davezies[†], Xavier D'Haultfoeuille[‡] and Denis Fougère[§]

First version: September 2006
Final version: July 2009

Abstract

This paper studies the econometric properties of a linear-in-means model of social interactions. Under a slightly more restrictive framework than Lee (2007), 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, for instance when endogenous and exogenous peer effects cancel each other. We extend this analysis to the case where only binary outcomes are observed. Once more, most parameters are semiparametrically identified under weak conditions. However, identifying all of them requires more stringent assumptions, including an homoskedasticity condition. We also develop a parametric estimator for the binary case which relies on the GHK simulator. Monte Carlo simulations illustrate the influence of group sizes on the accuracy of the estimation, in line with the results obtained by Lee (2007).

Keywords: social interactions; linear-in-means model; semiparametric identification.

*We would like to thank the editor, two anonymous referees, Stéphane Grégoir, Steve Machin, Amine Ouazad, as well as the participants in the CEPR Summer School on "The Economics of Education and Education Policy in Europe" (Padova) and in the CREST seminar (Paris) for their useful comments.

[†]CREST-INSEE, 18 boulevard Gabriel Péri 92245 Malakoff Cedex, France.
[‡]CREST-INSEE, 18 boulevard Gabriel Péri 92245 Malakoff Cedex, France.
[§]CNRS and CREST-INSEE, 18 boulevard Gabriel Péri 92245 Malakoff Cedex, France, CEPR, London, and IZA, Bonn.

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 identified under some strong exogeneity conditions. In the context of pupil achievement for instance, Hoxby (2000) and Ammermueller & Pischke (2006) reach identification by assuming that variations in time or across 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 group size variation.

The aim of our paper is threefold. Firstly, we reexamine the identification of this linear-in-means model when group sizes do not depend on the sample size.[2] 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, this extra assumption enables us to clarify the sources of identification in this model.[3] More precisely, we show that in the linear-in-means model, the crucial assumptions for identification are 1) the knowledge of the group sizes, and 2) the fact that group sizes take at least three different values. Parametric assumptions on the error term are not required. In general, homoskedasticity is not required either. This last assumption is useful however when both types of peer effects cancel each other, since in this case identification is lost without such a restriction.

Secondly, we extend these results to a model where only binary outcomes are observed. Identification of discrete outcome models with social interactions has already been studied by, e.g., Brock & Durlauf (2001, 2007) and Krauth (2006). 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. This model is close to spatial discrete choice models (see e.g. Case, 1992, McMillen, 1992, Pinkse & Slade, 1998, Beron & Vijverberg, 2004 or Klier & McMillen, 2008). The difference is that we allow here for

[1] Subsequently, we will often consider the example of peer effects in schools, 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 (Duflo0 & Saez, 2003).

[2] This is approximately the scenario with small group interactions proposed by Lee (2007).

[3] Under his more general setting, Lee (2007) provides sufficient conditions for identification, but they are rather difficult to interpret (see his Assumption 6.1 and 6.2).

exogenous peer effects and for fixed group effects simultaneously. The attractive feature of our result is that it does not rely on any functional assumption concerning the errors. Once more, the exogenous peer effects can be identified through group size variation. On the other hand, due to the loss of information, endogenous peer effects cannot be identified without further restrictions. We show that an homoskedasticity condition is sufficient for this purpose.

Thirdly, we develop a parametric estimation of the binary model, complementing the methods proposed by Lee (2007) for the model with a continuous outcome. We show that under a normality assumption on the residuals and a linear specification à la Mundlak (1961) on the fixed effect, a simulated maximum likelihood estimator can be implemented by using the GHK algorithm (Geweke, 1989, Keane, 1994 and Hajivassiliou et al., 1996). Thus, this estimator is close to Beron and Vijverberg (2004)'s one on spatial probit models. We investigate its finite sample properties through Monte Carlo simulations. The results stress the determining effect of average group size for the accuracy of the inference, in line with Lee (2007)'s result concerning the linear model.

The paper is organized as follows. In the next section, we present the theoretical model of social interactions. In Section 3, we study the identification of the model, both for the continuous and the discrete cases. The fourth section discusses the parametric estimation method of the discrete model. Section 5 displays Monte Carlo simulations. Section 6 concludes. Proofs are given in the appendix.

2 A theoretical model of social interactions

We consider the issue of individual choices in the presence of social interactions within groups. Let e_i denote the continuous choice variable of an individual i who belongs to a group of size m, x_i be her exogenous covariates and ε_i her (random) individual-specific characteristic. We suppose that her utility when choosing e_i, while the other persons in the group choose $(e_j)_{j\neq i}$, takes the following form:

$$\begin{aligned}\mathcal{U}_i(e_i,(e_j)_{j\neq i}) \;&=\; e_i\Bigg[x_i\beta_{10}+\left(\frac{1}{m-1}\sum_{j=1,\,j\neq i}^{m} e_j\right)\lambda_0\\&\qquad+\left(\frac{1}{m-1}\sum_{j=1,\,j\neq i}^{m} x_j\right)\beta_{20}+\alpha+\varepsilon_i\Bigg]-\frac{1}{2}e_i^2.\end{aligned}$$

In this framework, the marginal returns of individual i depends on her own characteristics x_i, her peers' choices $(e_j)_{j\neq i}$, their observable (exogenous) characteristics $(x_j)_{j\neq i}$ and a group fixed effect α. In a classroom, for instance, the utility of a student depends on her effort e_i and on the efforts of others because of spillovers in the learning process. Like Cooley (2007) and Calvó-Armengol et al. (2008), among others, we also allow the utility of each individual to depend on the observable characteristics of her peers. Indeed, there is some empirical evidence about the influence of peers' race, gender or parental education on student achievement (see e.g. Hoxby, 2000, or Cooley, 2007). A plausible explanation is that the marginal effect of e_i on achievement (which is positively correlated with the student's utility) depends on these characteristics.[4] Lastly, the outcome may depend on a classroom-specific effect, because of the teacher's quality, for instance. This model is close to the one considered by Calvó-Armengol et al. (2008) who study the effect of peers on individual achievement at school. An important difference is that they consider the network of friends, whereas our model is better suited when all classmates potentially affect the student's achievement.

Assuming that α and the $(x_i, \varepsilon_i)_{1 \leq i \leq m}$ are observed by all the individuals in the group, the Nash equilibrium of the game $(\widetilde{y}_1, ..., \widetilde{y}_m)$ satisfies

$$\widetilde{y}_i = x_i \beta_{10} + \left(\frac{1}{m-1} \sum_{j=1, j \neq i}^{m} \widetilde{y}_j \right) \lambda_0 + \left(\frac{1}{m-1} \sum_{j=1, j \neq i}^{m} x_j \right) \beta_{20} + \alpha + \varepsilon_i. \tag{2.1}$$

This model is identical to Lee's model (2007) of social interactions. 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 effect and α is a contextual (group-specific) effect. This model departs from the one considered by Manski (1993) or by Graham & 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.[5] Interestingly, one can show that Manski's model is actually the Bayesian Nash equilibrium of the game when player i does not observe the characteristics $(x_j, \varepsilon_j)_{j \neq i}$ of her peers, the $(\varepsilon_i)_{1 \leq i \leq m}$ being mutually independent and independent of $(x_1, ..., x_m, \alpha, m)$. This framework seems more realistic in large groups, whereas the hypothesis that the characteristics of other persons in the group are observed is likely to hold in small ones.

[4] For instance, girls may less disrupt classrooms than boys, all things being equal (see Lazear, 2001). In this case, the marginal effect of effort increases for everyone in the classroom.

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

3 Identification

We now turn to the identification of Model (2.1). First, as a benchmark, we suppose that the outcomes $\widetilde{y}_i$ are directly observed. This case corresponds to Lee (2007)'s framework, but we investigate it under a slightly different approach in Subsection 3.1. In Subsection 3.2, we study the situation where only rough measures of the outcomes, namely $y_i = \mathbb{1}\{\widetilde{y}_i \geq 0\}$, are available. In both cases, we implicitly assume that the econometrician knows the group of interactions for each individual. In the previous example, this assumption is mild if students really interact within the classroom, since the classroom identifier is usually known. It can be restrictive otherwise, but at our best knowledge, this assumption is also maintained in all papers studying identification of peer effects, including those by Manski (1993), Brock and Durlauf (2001), Lee (2007), Graham (2008) and Bramoullé et al. (2009). This stems from the fact that, in Manski's model at least, very little can be inferred from the data and from the model if the peer group is not known (see Manski, 1993, Subsection 2.5).

3.1 The benchmark: the linear model

In this section, we clarify the results obtained by Lee (2007), in the case where the size m of the group does not depend on the size of the sample.[6] We believe that such an assumption is virtually always satisfied in practice. For instance, there is no reason why the mean classroom size should depend on the size of the sample. Moreover, this restriction enables us to show what is identified from the usual exogeneity condition (see Assumption 4 below) and when homoskedasticity is required (see Theorem 3.2 below).

It is quite common to observe some but not all members in each group, and we take this into account for identification. On the other hand, we maintain the assumption that the size of the group is observed.[7] Let n denote the number of sampled individuals in the group ($n \leq m$). We denote by $\widetilde{Y}$ (respectively, X) the vector of outcomes $\widetilde{y}_i$ (respectively, of covariates) of the individuals sampled in the group. Let $F_{m,n}$ denote the distribution function of (m, n) and $F_{\widetilde{Y},X|m,n}$ denote the conditional distribution of $(\widetilde{Y}, X)$ given (m, n). Lastly, we denote by $\text{Supp}(T)$ the support of a random variable T. We rely on the following definition of identification.

[6] This is approximately the scenario with small group interactions considered by Lee (2007).

[7] This assumption is realistic in our leading example. In French panels of students, for instance, classroom sizes are observed while only a fraction of pupils within classrooms is sampled.

Definition 1 $(\beta_{10}, \beta_{20}, \lambda_0)$ *is identified if there exists a function* φ *such that*

$$(\beta_{10}, \beta_{20}, \lambda_0) = \varphi \left[\left(F_{\widetilde{Y}, X | m=m^*, n=n^*} \right)_{(m^*, n^*) \in \text{Supp}(m,n)}, F_{m,n} \right].$$

This definition states that the structural parameters are identified if they can be obtained through the distribution of the data. Implicit in the definition is the fact that our asymptotic is in the number of groups, as it is the case in standard panel data models.[8] Now, the key point for identification of the parameters when the $\widetilde{y}_i$ are observed is to focus on the within-group equation, which may be written as:

$$W_n \widetilde{Y} = W_n X \left(\frac{(m-1)\beta_{10} - \beta_{20}}{m - 1 + \lambda_0} \right) + W_n \frac{U}{1 + \lambda_0/(m-1)}, \tag{3.1}$$

where U is the vector of unobserved residuals ε for individuals sampled in the group, and W_n denotes the within-group matrix of size n, that is to say the matrix with $(1 - 1/n)$ on the diagonal and $(-1/n)$ elsewhere. To identify 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:

Assumption 1 $\Pr(n \geq 2) > 0$.

Assumption 2 $\text{Supp}(m)$ *contains at least three values.*

Assumption 3 *For all* $1 \leq i, j \leq m$, $E\left[x_i' \varepsilon_j \mid m, n\right] = 0$.

Assumption 4 $E\left[X' \ W_n \ X \mid m, n\right]$ *is almost surely nonsingular.*

Assumption 5 $1 > \lambda_0 > 1 - \min\left(\text{Supp}(m)\right)$.

Assumption 1 simply states that the within-group approach is feasible. Assumption 2, which is the cornerstone of our approach, ensures that there is sufficient variation in group sizes. Assumptions 1, 3 and 4 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 does not cause any trouble if, for instance, the observed individuals are drawn randomly in each group. Finally, Assumption 5 ensures that $\beta(m^*)$ exists for all $m^* \in \text{Supp}(m)$.[9]

[8] Indeed, when the number of groups tend to infinity, we are able to estimate consistently $\left(F_{\widetilde{Y}, X | m=m^*, n=n^*} \right)_{(m^*, n^*) \in \text{Supp}(m,n)}$ as well as $F_{m,n}$.

[9] Theorem 3.1 would remain valid if Assumption 5 were replaced by the weaker condition $\lambda_0 \notin -\text{Supp}(m-1)$. However, Assumption 5 is required under this form in Theorems 2, 3, 4 and in Lemma 1.

Theorem 3.1 *Under Assumptions 1-5, 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 only up to scale.

Theorem 3.1 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 $\beta_{20}' = -\lambda_0'\beta_{10}$. Using the first conditional moment of $\widetilde{Y}$ alone, one cannot distinguish the case with both exogenous and endogenous peer effects (which cancel out in this case) from the case with no peer effects. 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 $\beta(.)$ is constant or not.

Contrary to the reduced form approach, we do not need to know the mean $(\overline{x}_r)_{1\leq r\leq R}$ in each group to identify the parameters. Thus the problem of measurement error on $\overline{x}_r$, 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 way. The issue of identification in this case is left for future research.[10]

The nature of the group size effect provides another identifying assumption. Indeed, m may be correlated with α in a general way, but we cannot add interaction terms between the indicators $\mathbb{1}\{m = m^*\}$ (with $m^* \in \text{Supp}(m)$) and the covariates to the list of regressors, since then Assumption 4 would fail. To see this, let us remark that, if β_{10} and β_{20} depend on m in an unspecified way, then we can still identify $\beta(m)$ but not the structural parameters. On the other hand, identification of these structural parameters can be achieved if the dependence of β_{10} and β_{20} with respect to m takes a parametric form.[11] Of course, in this case, identification requires that m takes more than three different values. This also implies that the basic model where β_{10}, β_{20} and λ_0 are constant across group sizes is overidentified as soon as we observe at least four different group sizes. A simple way to test this restriction is to estimate $\beta(.)$ by using a within-group estimator for each group size, and then to implement the overidentification test for minimum distance estimators

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

[11]For instance we can write these parameters as affine transformations of m. This is equivalent to adding interaction terms between X and m.

(see e.g. Wooldridge, 2002, p. 444).

If $\beta_{20} = -\lambda_0\beta_{10}$, then λ_0 and β_{20} cannot be identified without further restriction. To recover them, one can use the residual variance variation, under an homoskedasticity condition (see our Assumption 6 below). More precisely, the conditional variance of the residuals should not depend on the group size. This hypothesis is quite weak since it does not restrict the relationship between the residuals ε_{ri} and the covariates x_{ri}. Moreover, under Assumption 6, one needs less variation across group sizes than previously, and we can replace Assumption 2 by Assumption 2'.

Assumption 2' Supp(m) *contains at least two values.*

Assumption 6 $V(U \mid n, m) = \sigma^2 I_n$ *where I_n is the identity matrix of size n.*

Theorem 3.2 *Under Assumptions 1, 2' and 3-6, $(\beta_{10}, \lambda_0, \beta_{20})$ are identified.*

The idea of using second order moments to identify peer effects has already been exploited by Glaeser et al. (1996) and Graham (2008). In particular, Graham (2008) develops a framework where composite peer effects can be identified through such a restriction. In his model, however, endogenous peer effects are not identified.

3.2 The binary model

We now investigate whether the parameters are still identified when one cannot observe directly the outcome variable $\widetilde{y}_i$ but only a rough binary measure of it, namely $y_i = \mathbb{1}\{\widetilde{y}_i \geq 0\}$.[12] For instance, when studying peer effects in the classroom, the analyst could observe only grade retention decisions rather than students' efforts. Similarly, in criminal studies, the violence level chosen by an individual may depend on the violence level chosen by her peers. The level chosen in equilibrium is a continuous variable. However, the econometrician may only be able to observe a rough measure of this violence level, through criminal acts. This fits within our framework as long as doing criminal acts corresponds to being above a given threshold of violence.

The binary model we consider is not a discrete choice model but rather a continuous choice model with imperfect observations of the choice. In discrete choice models, the econometrician observes the choice $\widetilde{y}_i \in \{1, ..., p\}$ of i. This choice depends on $(\widetilde{y}_j)_{j\neq i}$,

[12] The definition of identification that we use here is similar to the one introduced in Definition 1, except that $\widetilde{Y}$ has to be replaced by Y, the vector of outcomes y_i observed for the individuals sampled in the group.

as in Equation (2.1), but in a nonlinear way. Such models have been studied by Brock and Durlauf (2001, 2007), Tamer (2003), Krauth (2006) and Bayer & Timmins (2007). The main challenge when making inference on these models is that in general, multiple equilibria arise. This is not a concern here, as $\widetilde{y}_i$ is uniquely defined by Equation (2.1).

When the outcome is a binary variable, the reduced-form Equation (3.1) is useless for identification since $W_n\widetilde{Y}$ has no observational counterpart. Instead, we rely on Equation (3.2) below.

Lemma 1 *Suppose that $y_i = \mathbb{1}\{\widetilde{y}_i \geq 0\}$, where $\widetilde{y}_i$ satisfies Equation (2.1), and that Assumption 5 holds. Then the model is observationally equivalent to the model generated by the following equation:*

$$\begin{aligned} y_i \quad = \quad & 1\Bigg\{x_i\left(\beta_{10} - \frac{\beta_{20}}{m-1}\right) + \left[\overline{x}\frac{m}{m-1}\left(\beta_{20} + \frac{\beta_{10}+\beta_{20}}{1-\lambda_0}\lambda_0\right) + \alpha\left(1+\lambda_0(m)\right)\right] \\ & +\overline{\varepsilon}\,\lambda_0(m) + \varepsilon_i \geq 0\Bigg\}, \end{aligned} \tag{3.2}$$

where $\lambda_0(m) = m\lambda_0/((m-1)(1-\lambda_0))$.

The term into brackets is a fixed group 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. In the following, we denote by x_j^k the k-th covariate of individual j. The following assumptions are needed for identification.[13]

Assumption 7 $(\varepsilon_1, ..., \varepsilon_m)$ *are exchangeable conditional on* $(m, x_1, ..., x_m, \alpha)$. *The support of* $\varepsilon_1 + \lambda_0(m)\overline{\varepsilon}$ *conditional on* $(m, x_1, ..., x_m, \alpha)$ *is* $\mathbb{R}$*, almost surely.*

Assumption 8 *Let $z = x_2 - x_1$. The support of z is not contained in any proper linear subspace of $\mathbb{R}^K$, where K denotes the dimension of x_i.*

Assumption 9 *There exists k_0 such that z^{k_0} has everywhere a positive Lebesgue conditional density given $(m, z^1, ..., z^{k_0-1}, z^{k_0+1}, ..., z^K)$ and such that $\beta_{10}^{k_0} = 1$. Without loss of generality, we set $k_0 = 1$.*

The first part of Assumption 7 holds for instance if, conditional on m and α, the residuals $(\varepsilon_i)_{1\leq i\leq m}$ are exchangeable and independent of the covariates $(x_i)_{1\leq i\leq m}$. In particular, Assumption 7 is satisfied if the $(\varepsilon_i)_{1\leq i\leq m}$ are i.i.d. and independent of $(x_1, ..., x_m, m, \alpha)$. The second part of Assumption 7 is a technical condition, which is identical to the second

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

part of Assumption 1 set forth by Manski (1987). Assumption 8 ensures that z varies enough within a group. As usually in binary models, one parameter must be normalized, and this is the purpose of Assumption 9. However, a small difficulty arises here, because the reduced form does not allow us to identify the sign of the structural parameters. A sufficient condition is to fix one parameter (and not only its absolute value): thus we set $\beta^1_{10} = 1$.[14]

Theorem 3.3 *Suppose that Assumptions 1-2, 5 and 7-9 hold. Then β_{10} is identified. Moreover,*
- if $\beta_{20} \neq \beta^1_{20}\beta_{10}$, then β_{20} is identified,
- if $\beta_{20} = \beta^1_{20}\beta_{10}$, β^1_{20} is not identified and the other parameters β^k_{20} are identified up to β^1_{20}.
On the other hand, λ_0 is not identified.

If fewer parameters (i.e., fewer than those included in Model (2.1) are identified, Theorem 3.3 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 set forth in Theorem 3.1, identification of β_{20} is lost when there is no exogenous peer effect, because in this case $\beta_{20} = \beta^1_{20}\beta_{10} = 0$. The non-identifiability of λ_0 is not surprising since this parameter only appears in the fixed effect and in the residuals (see Equation (3.2)). Heuristically, without any assumption imposed on these terms, any λ_0 can be rationalized by changing accordingly α and the residuals $(\varepsilon_i)_{1\leq i\leq m}$.

Thus stronger assumptions are needed for identifying λ_0. One possibility is to observe $\overline{x}$ and to restrict the dependence between the residuals and the covariates through the following assumptions:

Assumption 2" *The support of m given $\overline{x}$ has at least three elements with positive probability.*

Assumption 10 *$\overline{x}$ is observed.*

Assumption 11 $(\varepsilon_1, ..., \varepsilon_m, \alpha) \perp\!\!\!\perp (x_1, ..., x_m) \mid m, \overline{x}$.

Assumption 12 $V(\varepsilon_1, ..., \varepsilon_m, \alpha \mid \overline{x}, m) = \begin{pmatrix} V(\varepsilon_1 \mid \overline{x})I_m & 0 \\ 0 & V(\alpha \mid \overline{x}) \end{pmatrix}$.

Assumption 13 *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 $\mathbb{R}^2$.*

[14]Obviously, Theorem 3.3 also holds with $\beta^1_{10} = -1$.

Assumption 2'' is slightly more restrictive than Assumption 2, but should hold most of the time. For instance, it is satisfied for a multinomial logit (or probit) model generating the conditional distribution of m given $\overline{x}$. As mentioned above, Assumption 10 is a restrictive condition as it imposes either to observe all individuals in the group or to consider only covariates whose means are known. Assumption 11 is in the same spirit as Assumption 7. It restricts the dependence between α and the covariates to a dependence on the mean. Assumption 12 is the assumption of homoskedasticity in m; it is very similar to Assumption 6. The difference between both assumptions stems from the identifying equation we use in both cases. In the discrete model, α remains in Equation (3.2), and thus its variance must be modeled as well as its covariance with the residuals $(\varepsilon_i)_{1\leq i\leq m}$.[15] Finally, Assumption 13 is a condition of large support. In particular, it implies that $m \geq 3$. Otherwise, indeed, the two variables belong to a line in $\mathbb{R}^2$.

Theorem 3.4 *Under Assumptions 1, 2'', 5, and 7-13, and if* $\beta_{20} \neq \beta_{20}^1\beta_{10}$*,* λ_0 *is also identified.*

4 Estimation

In this section we restrict the analysis to the case where only $\mathbb{1}\{\widetilde{y}_i \geq 0\}$ is observed, since the continuous case is analyzed in full details by Lee (2007). We also restrict ourselves to a parametric setting with homoskedasticity that is characterized by the following assumptions:

Assumption 14 *The residuals* $(\varepsilon_i)_{1\leq i\leq m}$ *are i.i.d. and* $\varepsilon_i \sim \mathcal{N}(0,1)$*.*

Assumption 15 $\alpha|\overline{x}, m \sim \mathcal{N}(\gamma_0(m) + \delta_0(m)\overline{x}, \sigma_0^2)$.

Assumption 14 imposes the normality of the residuals. This assumption is also imposed by Lee (2007) when he develops his conditional maximum likelihood estimator, or by McMillen (1992) and Beron and Vijverberg (2004), among others, when studying spatially dependent discrete choice models. Contrary to the previous section, we adopt here the usual normalization by supposing that the variance of the residuals is equal to one. Assumption 15 has two consequences. First, it strengthens Assumptions 11 and 12 by introducing a linear dependence à la Mundlak (1961) between α and $\overline{x}$, conditional on m. Note that the

[15] The assumption of no covariance is not restrictive. Indeed, if there is a correlation between ε_i and α which does not depend on i, one can always reparametrize the model in order to make them uncorrelated.

dependence between α and m remains very flexible. Second, Assumption 15 imposes the normality of the residual term, in a similar way to the standard random effect probit.

Under these conditions, the model is fully identified, as in Theorem 3.4 but in a more direct way. Indeed, β_{10} and β_{20} can be identified through group size variations. Moreover, the model can be written in this case as

$$y_i = 1\left\{\gamma_0'(m) + x_i\left(\beta_{10} - \frac{\beta_{20}}{m-1}\right) + \overline{x}\,\delta_0'(m) - v_i \geq 0\right\}, \tag{4.1}$$

where $\gamma_0'(m)$ and $\delta_0'(m)$ depend on $\gamma_0(m)$, $\delta_0(m)$ and on the parameters of the model, the error term v_i being a combination of $(\varepsilon_i)_{1\leq i\leq m}$ with the residual $\alpha - \gamma_0(m) - \delta_0(m)\overline{x}$. Conditional on m, the vector $(v_i)_{1\leq i\leq m}$ is normally distributed and exchangeable, with

$$\begin{aligned} V(v_i|m) &= 1 + \sigma_0^2 + \lambda_0(m)(2+\lambda_0(m))\left(\sigma_0^2 + \frac{1}{m}\right), \\ \mathrm{Cov}(v_i, v_j|m) &= \sigma_0^2 + \lambda_0(m)(2+\lambda_0(m))\left(\sigma_0^2 + \frac{1}{m}\right), \qquad \forall\, i \neq j. \end{aligned}$$

One can show that when m varies, it is possible to separate λ_0 from σ_0^2 in the covariances (or in the variance).

Now, let us suppose that we observe a sample of R groups where, for the sake of simplicity, all members in each group are observed (even if we only need to observe $\overline{x}$). Hence, for group r, we observe its size m_r, the vector of outcomes $Y_r = (y_{r1}, ..., y_{rm_r})$ and the vector of covariates $X_r = (x_{r1}, ..., x_{rm_r})$. We suppose that the sizes $(m_r)_{1\leq r\leq R}$ are i.i.d., and that $(X_r, \alpha_r, V_r)_{1\leq r\leq R}$ are independent and distributed according to $F_{X,\alpha,V|m,n}$, where V is the vector of unobserved shocks $(v_1, ..., v_m)$. In the previous example of peer effects in the classroom, this condition imposes that there is no spillovers between classrooms.

Let $\theta = (\beta_1, \beta_2, \lambda, \sigma^2, (\gamma'(m^*), \delta'(m^*))_{m^*\in\mathrm{Supp}(m)})$ denote the vector of all parameters. Under the previous i.i.d. assumption, the likelihood of the whole sample satisfies

$$L(Y_1, ..., Y_R|m_1, ..., m_R, X_1, ...X_R, \theta) = \prod_{r=1}^{R} L(Y_r|m_r, X_r, \theta),$$

where $L(Y_r|m_r, X_r, \theta)$ denotes the likelihood for group r. Moreover, by using (4.1), we can write this likelihood as:

$$\begin{aligned} & L(Y_r|m_r, X_r, \theta) \\ = \ & \Pr\left[(2y_{r1}-1)v_{r1} \leq (2y_{r1}-1)\left(\gamma'(m_r) + x_{r1}\left(\beta_1 - \frac{\beta_2}{m_r-1}\right) + \overline{x}_r\delta'(m_r)\right), ..., \right. \\ & \left. (2y_{rm_r}-1)v_{rm_r} \leq (2y_{rm_r}-1)\left(\gamma'(m_r) + x_{rm_r}\left(\beta_1 - \frac{\beta_2}{m_r-1}\right) + \overline{x}_r\delta'(m_r)\right)\right]. \end{aligned}$$

This is the probability that a multivariate normal vector belongs to an hyperrectangle in $\mathbb{R}^{m_r}$. Such a probability can be estimated, for instance, by the GHK algorithm (Geweke, 1989, Keane, 1994, and Hajivassiliou et al., 1996). Thus, the model can be estimated by simulated maximum likelihood.

5 Monte Carlo simulations

In this section, we investigate the finite sample performance of our estimator. The sample data are generated with one regressor $x_{ri} \sim \mathcal{N}(0,4)$, the $(x_{ri})_{r,i}$ being independent for all r and i. The true parameters are $\beta_{10} = 1$, $\beta_{20} = 1$, $\lambda = 0.2$, $\sigma_0^2 = 0.5$, $\gamma(m) = 0$ for all m, and $\delta(m) = 0.1$ for all m. As Lee (2007), we consider a case where the average size group is small, and another where it is relatively large. In the first case, the group sizes vary from 3 to 8, the number of groups of each size being the same. In the relatively large case, they range from 15 to 25. The first case could be realistic for groups of good friends or roommates for instance, whereas the second one could correspond to groups of students in a classroom. In each case, we consider different sample sizes from 330 to 21,120. In the GHK algorithm, we use Halton sequences instead of standard uniform random numbers as they improve, on average, the accuracy of the integral estimation (see, e.g., Sándor & András, 2004). In the small group case where the dimension of the integral is low, we rely on 25 replications, whereas we utilize 50 replications in the large group case.

Table 1 displays our results. The first striking point is that sample sizes must be quite large to obtain satisfactory results. If we compare the results of our small groups scenario with the one considered by Lee (see Lee, 2007, Table 1, Model SG-SX), it seems that, observing a binary measure of $\widetilde{y}_i$ instead of $\widetilde{y}_i$ itself leads to rather large biases for even moderately large sample sizes.[16] In particular, the bias on λ_0 is systematically negative for small and moderately large sample sizes. The second striking result is the influence of the group sizes. The accuracy of the estimator of β_{20} in large groups is approximately the same as the one in small groups, but with a sample four times larger. This is not surprising, since identification of peer effects becomes weak as the sample size increases (see Lee, 2007). The parameter λ_0 is also better estimated with small groups, but the difference between the two designs seems to decrease when the sample size grows. On the other hand, and quite surprisingly, the estimator of β_{10} is more precise in large groups.

[16]Note that it is difficult to compare our large group scenario with the one studied by Lee, since he considers a model with two independent covariates x_{1i} and x_{2i} such that x_{1i} has only a direct effect on y_i (i.e., $\beta_{20}^1 = 0$), while x_{2i} affects y_i only through exogenous peer effects (so that $\beta_{10}^2 = 0$).

Table 1: Results of the Monte Carlo simulations.

Sample size	Parameter	Small groups Mean	Small groups Std. err.	Large groups Mean	Large groups Std. err.
660	β_{10}	0.9975	0.2254	1.0128	0.1658
	β_{20}	0.8956	0.8877	1.4445	2.7601
	λ_0	-0.0304	0.5688	-0.3801	0.6600
1320	β_{10}	1.0029	0.1198	1.0025	0.0865
	β_{20}	0.9823	0.4885	0.9780	1.4712
	λ_0	0.1158	0.3458	-0.0026	0.3093
2640	β_{10}	0.9936	0.0951	0.9978	0.0678
	β_{20}	0.9378	0.3739	1.0761	0.8625
	λ_0	0.1831	0.1405	0.1247	0.1833
5280	β_{10}	0.9904	0.0664	1.0001	0.0419
	β_{20}	0.9744	0.2425	1.0264	0.5747
	λ_0	0.1927	0.0678	0.1620	0.1167
10560	β_{10}	0.9914	0.0451	1.0014	0.0285
	β_{20}	0.9708	0.1690	1.0240	0.4303
	λ_0	0.2000	0.0389	0.1788	0.0513
21120	β_{10}	0.9911	0.0295	0.9984	0.0180
	β_{20}	0.9872	0.1065	0.9777	0.2847
	λ_0	0.1897	0.0284	0.1950	0.0311

Note: The small groups scenario corresponds to a sample composed of groups whose size goes from 3 to 8, the number of groups of different sizes being equal. The large groups scenario corresponds to a sample of groups whose size goes from 15 to 25, the number of groups of different sizes being still equal.

6 Conclusion

This paper considers identification and estimation of social interaction models using group size variation. Provided that the sizes of the groups are known and vary sufficiently, endogenous and exogenous peer effects can be identified without any exclusion restriction in the linear-in-means model. The result can be extended to a binary outcome model. In this case, exogenous peer effects are also identified under weak assumptions. Identification of endogenous peer effects is more stringent, as it requires an homoskedasticity condition and restrictions on the dependence between fixed group effects and covariates.

Our paper has two main limitations. First, the size of each group is assumed to be known. However, as emphasized by Manski (2000), it is often difficult to define groups a priori. 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 should be examined in a future research.

Appendix A: proofs

In all the proofs, t^* denotes a possible value of the random variable t.

Proof of Theorem 3.1: First, under Assumption 3, $E(X'W_nU \mid n, m) = 0$. Thus, by Assumption 4, $\beta(m^*)$ is identified for all $m^* \in \text{Supp}(m)$. We now prove that the knowledge of $m^* \mapsto \beta(m^*)$ allows in general to identify the structural parameters.

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

$$(\beta_{10}\lambda_0 + \beta_{20})(m_1^* - m_2^*) = 0.$$

Hence, if $\beta_{20} = -\lambda_0\beta_{10}$, $\beta(.)$ is constant. In the opposite case, $\beta(.)$ 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 identified by $\beta(.)$. Since $\beta_{20} = -\lambda_0\beta_{10}$, β_{20} is identified up to a constant.

Now suppose that $\beta_{20} \neq -\lambda_0\beta_{10}$. Let (m_0^*, m_1^*, m_2^*) be three different values in $\text{Supp}(m)$. We will prove that the knowledge of $\beta(m_0^*), \beta(m_1^*)$ and $\beta(m_2^*)$ allows to identify $(\beta_{10}, \lambda_0, \beta_{20})$. 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

$\text{com}(A)'B = 0$ where $\text{com}(A)$ 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,

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

Thus,

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

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.

Proof of Theorem 3.2: Because $m^* \mapsto \beta(m^*)$ is identified, $V\left(\frac{W_n U}{1+\frac{\lambda_0}{m-1}} \mid n, m\right)$ is known. Thus, under Assumption 6,

$$V\left(\frac{W_n U}{1 + \frac{\lambda_0}{m-1}} \mid n, m\right) = \frac{\sigma^2}{\left(1 + \frac{\lambda_0}{m-1}\right)^2} W_n.$$

Hence, for all $m_1^* \neq m_2^* \in \text{Supp}(m)^2$,

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

is identified. Under Assumption 5, $1 + \lambda_0/(m^* - 1) > 0$ for all $m^* \in \text{Supp}(m)$. 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 $m_1^* = m_2^*$, which is a contradiction. Thus λ_0 is identified.

Then, because $m^* \mapsto \beta(m^*)$ is identified, $\beta_{10} - \beta_{20}/(m^* - 1)$ is known for all $m^* \in \text{Supp}(m)$. Taking two different values for m^* allows to identify β_{20} and thus β_{10}.

Proof of Lemma 1: Taking the mean in both sides of Equation (2.1) leads to

$$\overline{\overline{y}} = \overline{x}\left(\frac{\beta_{10} + \beta_{20}}{1 - \lambda_0}\right) + \frac{\alpha}{1 - \lambda_0} + \frac{\overline{\varepsilon}}{1 - \lambda_0},$$

since $1/(1-\lambda_0)$ exists, according to Assumption 5. Because $\sum_{j\neq i}\widetilde{y}_j = m\overline{\widetilde{y}} - \widetilde{y}_i$ and $\sum_{j\neq i} x_j = m\overline{x} - x_i$, Equation (2.1) is then equivalent to

$$\begin{aligned}\widetilde{y}_i\left(1+\frac{\lambda_0}{m-1}\right) &= x_i\left(\beta_{10}-\frac{\beta_{20}}{m-1}\right)+\overline{x}\frac{m}{m-1}\left(\beta_{20}+\frac{\beta_{10}+\beta_{20}}{1-\lambda_0}\lambda_0\right)\\ &\quad+\alpha\left(1+\frac{m}{m-1}\frac{\lambda_0}{1-\lambda_0}\right)+\overline{\varepsilon}\frac{m}{m-1}\frac{\lambda_0}{1-\lambda_0}+\varepsilon_i.\end{aligned}$$

Now, under Assumption 5, $1+\lambda_0/(m-1)>0$, so that $\widetilde{y}_i \geq 0$ if and only if $\widetilde{y}_i(1+\lambda_0/(m-1))\geq 0$. Thus, under Assumption 5, $y_i = \mathbb{1}\{\widetilde{y}_i \geq 0\}$, where $\widetilde{y}_i$ satisfies Equation (2.1), is observationally equivalent to y_i satisfying Equation (3.2).

Proof of Theorem 3.3: Assumption 7 implies that the conditional distribution of $\varepsilon_i + \lambda_0(m)\overline{\varepsilon}$ is identical for every i. Thus Assumption 1 in Manski (1987) is satisfied and, using our Assumptions 8 and 9, we can apply directly Manski's result to identify $((m-1)\beta_{10} - \beta_{20})/|m-1-\beta_{20}^1|$. The first term of the vector, $((m-1)\beta_{10}^1 - \beta_{20}^1)/|m-1-\beta_{20}^1|$, is also identified. By Assumption 9,

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

so that $\widetilde{\beta}(m)$ is identified as the ratio of two known terms. The rest of the proof for the identification of (β_{10},β_{20}) follows the same development than the one used for Theorem 3.1, λ_0 being replaced by $-\beta_{20}^1$.

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 the parameters $(\lambda_0', \alpha', \varepsilon_1', ..., \varepsilon_m')$ are observationally equivalent to those characterizing the initial model. Indeed, we can check that they lead to Equation (3.2) 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 $(\varepsilon_1', ..., \varepsilon_m')$. Furthermore, letting $\lambda_0'(m) = m\lambda_0'/((m-1)(1-\lambda_0'))$, we get

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

where
$$\alpha^* = \frac{(m-1+\lambda_0')(1-\lambda_0)\alpha'^* - m\overline{x^*}(\beta_{10}+\beta_{20})(\lambda_0-\lambda_0')}{(m-1+\lambda_0)(1-\lambda_0')}$$
and $\overline{x^*} = (1/m)\sum_i x_i^*$. Thus the second part of Assumption 7 also holds with $(\lambda_0', \alpha', \varepsilon_1',$ $..., \varepsilon_m')$. This shows that λ_0 is not identified.

Proof of Theorem 3.4: Let $\theta_0 = \lambda_0/(1-\lambda_0)$ and
$$\nu_i = \left[\overline{x}\frac{m}{m-1}\left[\beta_{20}+\theta_0(\beta_{10}+\beta_{20})\right] + \alpha\left(1+\frac{m}{m-1}\theta_0\right)\right] + \overline{\varepsilon}\frac{m}{m-1}\theta_0 + \varepsilon_i.$$
Note that $F_{\nu_1,...,\nu_m|x_1,...,x_m,m} = F_{\nu_1,...,\nu_m|\overline{x},m}$. Indeed
$$\begin{aligned}
& F_{\nu_1,...,\nu_m|x_1,...,x_m,m}(\nu_1^*,...,\nu_m^*|x_1^*,...,x_m^*,m^*) \\
= & \int F_{\nu_1,...,\nu_m|x_1,...,x_m,m,\alpha}(\nu_1^*,...,\nu_m^*|x_1^*,...,x_m^*,m^*,\alpha^*)dF_{\alpha|x_1,...,x_m,m}(\alpha^*|x_1^*,...,x_m^*,m^*) \\
= & \int F_{\nu_1,...,\nu_m|\overline{x},\alpha,m}(\nu_1^*,...,\nu_m^*|\overline{x^*},\alpha^*,m^*)dF_{\alpha|\overline{x},m}(\alpha^*|\overline{x^*},m) \\
= & F_{\nu_1,...,\nu_m|\overline{x},m}\left(\nu_1^*,...,\nu_m^*|\overline{x^*},m^*\right),
\end{aligned}$$
where the third line is derived from Assumption 11 and the fact that, given $(x_1,...,x_m,$ $m,\alpha)$, $(\nu_1,...,\nu_m)$ is a deterministic function of $(\varepsilon_1,...,\varepsilon_m)$. Moreover,
$$\begin{aligned}
& \Pr(y_1=0,y_2=0 \mid x_1=x_1^*,x_2=x_2^*,\overline{x}=x,m=m^*) \\
& = \Pr\left\{\nu_1 \le -x_1^*\left(\beta_{10}-\frac{\beta_{20}}{m-1}\right),\ \nu_2 \le -x_2^*\left(\beta_{10}-\frac{\beta_{20}}{m-1}\right)|x_1=x_1^*,x_2=x_2^*,\overline{x}=x,\right. \\
& \qquad \left. m=m^*\right\} \\
& = F_{\nu_1,\nu_2|\overline{x},m}\left(-x_1^*\left(\beta_{10}-\frac{\beta_{20}}{m^*-1}\right), -x_2^*\left(\beta_{10}-\frac{\beta_{20}}{m^*-1}\right)|x,m^*\right).
\end{aligned}$$
Since Theorem 3.3 implies that (β_{10},β_{20}) is identified, $x_1^*\left(\beta_{10}-\beta_{20}/(m^*-1)\right)$ and $x_2^*\left(\beta_{10}-\beta_{20}/(m^*-1)\right)$ are known. Moreover, $\overline{x}$ is observed so that the first term is identified on the whole support of (x_1,x_2) conditional on $(\overline{x},m)$. Thus, by Assumption 13, making (x_1,x_2) vary allows us to identify the whole conditional distribution of (ν_1,ν_2) given $\overline{x}$ and m.

Now, by Assumption 12,
$$\text{Cov}(\nu_1,\nu_1-\nu_2 \mid \overline{x},m) = \text{Cov}\left(\overline{\varepsilon}\frac{m}{m-1}\theta_0+\varepsilon_1,\varepsilon_1-\varepsilon_2 \mid \overline{x},m\right) = V(\varepsilon_1 \mid \overline{x}),$$
so that the right-hand side term is identified. Moreover, a little algebra shows that
$$\begin{aligned}
(m-1)^2\text{Cov}(\nu_1,\nu_2|\overline{x},m) \;=\; & m^2\left[(1+\theta_0)^2V(\alpha|\overline{x})\right] + m\Big[-2(1+\theta_0)V(\alpha|\overline{x}) \\
& +\theta_0(2+\theta_0)V(\varepsilon_1|\overline{x})\Big] + \left[V(\alpha|\overline{x}) - 2\theta_0V(\varepsilon_1|\overline{x})\right].
\end{aligned}$$

Conditional on $\overline{x}$, this is a regression of the (known) left term on $(m^2, m, 1)$. By Assumption $2''$, there exists a set A of positive probability such that m can take three different values with positive probability, given that $\overline{x} = x^*$ for all $x^* \in A$. Thus, the coefficients (a, b, c) of this regression can be identified. These coefficients depend on $\overline{x}$ but, for the sake of simplicity, this dependence is let implicit in the following. We show now that the knowledge of these coefficients implies that θ_0 is identified. The conclusion follows since θ_0 is in a one-to-one relationship with λ_0.

First, we set $\phi_0 = 1 + \theta_0$ and $\rho_0 = V(\alpha|\overline{x})/V(\varepsilon_1|\overline{x})$. We also define $a' = a/V(\varepsilon_1|\overline{x})$, $b' = b/V(\varepsilon_1|\overline{x}) + 1$ and $c' = c/V(\varepsilon_1|\overline{x}) - 2$. Then a', b' and c' are identified, and

$$\begin{cases} \phi_0^2\rho_0 &= a' \\ -2\phi_0\rho_0 + \phi_0^2 &= b' \\ \rho_0 - 2\phi_0 &= c'. \end{cases}$$

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

$$\begin{cases} \phi_0^3 + \frac{c'}{2}\phi_0^2 - \frac{a'}{2} &= 0 \\ \phi_0^2 + \frac{2c'}{3}\phi_0 + \frac{b'}{3} &= 0 \\ \rho_0 - 2\phi_0 &= c'. \end{cases} \tag{A.1}$$

This system admits at most two solutions in terms of (ρ, ϕ). Suppose that there exist 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 that, for all $\phi \in \mathbb{R}$,

$$\phi^3 + \frac{c'}{2}\phi^2 - \frac{a'}{2} = \left(\phi^2 + \frac{2c'}{3}\phi^2 + \frac{b'}{3}\right)(\phi + x).$$

Thus, $x = -c'/6$ and $2c'x = -b'$, which implies that $c'^2 = 3b'$. Replacing b' and c' by their values in terms of ϕ_i and ρ_i $(i \in \{0, 1\})$, we obtain, for $i \in \{0, 1\}$,

$$3(-2\phi_i\rho_i + \phi_i^2) - (\rho_i - 2\phi_i)^2 = 0.$$

This equation is equivalent to $\phi_i + \rho_i = 0$. Replacing ρ_i by $-\phi_i$ in c' yields $\phi_0 = \phi_1 = -c'/3$ and thus also $\rho_0 = \rho_1$. This contradicts our assumption that $(\rho_0, \phi_0) \neq (\rho_1, \phi_1)$. Thus ϕ_0 is identified by System (A.1), and the conclusion follows.

References

Ammermueller, A. & Pischke, J. S. (2006), Peer effects in european primary schools: Evidence from pirls. IZA Discussion Paper No. 2077.

Bayer, P. & Timmins, C. (2007), 'Estimating equilibrium models of sorting across location', *Economic Journal* **117**, 353–374.

Beron, K. J. & Vijverberg, W. P. M. (2004), Probit in a spatial context: A monte carlo analysis, *in* L. Anselin, R. J. G. M. Florax & S. J. Rey, eds, 'Advances in Spatial Econometrics: Methodology, Tools and Applications', Berlin: Springer-Verlag, pp. 169–196.

Bramoullé, Y., Djebbari, H. & Fortin, B. (2009), 'Identification of peer effects through social networks', *Journal of econometrics* **50**, 41–55.

Brock, W. A. & Durlauf, S. M. (2001), 'Discrete choice with social interactions', *Review of Economic Studies* **68**, 235–260.

Brock, W. A. & Durlauf, S. M. (2007), 'Identification of binary choice models with social interactions', *Journal of Econometrics* **140**, 52–75.

Calvó-Armengol, J., Pattacchini, E. & Zenou, Y. (2008), 'Peer effects and social networks in education', Forthcoming in *Review of Economic Studies* .

Case, A. (1992), 'Neighborhood influence and technological change', *Regional Science and Urban Economics* **22**, 491–508.

Cooley, J. (2007), Desegregation and the achievement gap: Do diverse peers help? Working Paper, Duke University.

Duflo0, E. & Saez, E. (2003), 'The role of information and social interactions in retirement plan decisions: Evidence from a randomized experiment', *Quarterly Journal of Economics* **118**, 815–842.

Geweke, J. (1989), 'Bayesian inference in econometric models using monte carlo integration', *Econometrica* **57**, 1317–1339.

Graham, B. S. (2008), 'Identifying social interactions through excess variance contrasts', *Econometrica* **76**, 643–660.

Graham, B. S. & Hahn, J. (2005), 'Identification and estimation of the linear-in-means model of social interactions', *Economics Letters* **88**, 1–6.

Hajivassiliou, V. A., McFadden, D. & Ruud, P. (1996), 'Simulation of multivariate normal rectangle probabilities and their derivatives: theoretical and computational results', *Journal of Econometrics* **72**, 85–134.

Hoxby, C. (2000), Peer effects in the classroom: Learning from gender and race variation. NBER Working Paper No 7867.

Keane, M. (1994), 'A computationally practical simulation estimator for panel data', *Econometrica* **62**, 95–116.

Klier, T. & McMillen, D. P. (2008), 'Clustering of auto supplier plants in the united states: Generalized method of moments spatial logit for large samples', *Journal of Business and Economic Statistics* **26**, 460–471.

Krauth, B. V. (2006), 'Simulation-based estimation of peer effects', *Journal of Econometrics* **133**, 243–271.

Lazear, E. (2001), 'Educational production', *Quarterly Journal of Economics* **116**, 777–803.

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.

Manski, C. F. (1987), 'Semiparametric analysis of random effects linear models from binary panel data', *Econometrica* **55**, 357–362.

Manski, C. F. (1993), 'Identification of endogenous social effects the reflection problem', *Review of Economic Studies* **60**, 531–542.

Manski, C. F. (2000), 'Economic analysis of social interactions', *Journal of Economic Perspectives* **14**, 115–136.

McMillen, D. P. (1992), 'Probit with spatial autocorrelation', *Journal of Regional Science* **32**, 335–348.

Mundlak, Y. (1961), 'Empirical production function free of management bias', *Journal of Farm Economics* **43**, 44–56.

Pinkse, J. & Slade, M. E. (1998), 'Contracting in space: An application of spatial statistics to discrete- choice models', *Journal of Econometrics* **85**, 125–154.

Rees, D. I., Zax, J. S. & Herries, J. (2003), 'Interdependence in worker productivity', *Journal of Applied Econometrics* **18**, 585–604.

Sándor, Z. & András, P. (2004), 'Alternative sampling methods for estimating multivariate normal probabilities', *Journal of Econometrics* **120**, 207–234.

Schennach, S. M. (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.

Wooldridge, J. M. (2002), *The Econometrics of Cross Section and Panel Data*, MIT Press.