Available online at www.sciencedirect.com
SCIENCE DIRECT®
Economics Letters 88 (2005) 1–6
economics letters
www.elsevier.com/locate/econbase

Identification and estimation of the linear-in-means model of social interactions

Bryan S. Graham[b], Jinyong Hahn[a,*]

[a]*Department of Economics, UCLA, Box 951477, Los Angeles, CA 90095-1477, United States*
[b]*Harvard University, United States*

Received 17 November 2004; accepted 2 February 2005

Abstract

This paper studies identification and estimation of the linear-in-means model of social interactions. Using a quasi-panel data approach, we show how endogenous social effects can be identified in the presence of unobserved group effects.

Keywords: Linear-in-means model of social interactions; Quasi-panel data approach

JEL classification: C1

1. Introduction

The idea that social context matters for individual outcomes forms the basis of a large and growing literature in economics. Peer or social group attributes and behavior are posited to affect individual behavior and vice versa. Much of this research is motivated by the observation that many individual outcomes, such as earnings, academic achievement, criminal behavior, sickness and unemployment, vary much more between social groups than within them. In attempting to empirically model social interactions, researchers typically augment regressions that include

* Corresponding author. Tel.: +1 310 825 2523; fax: +1 310 825 9528.
E-mail address: hahn@econ.ucla.edu (J. Hahn).

0165-1765/$ - see front matter
doi:10.1016/j.econlet.2005.02.001

individual-level covariates with measures of group-level characteristics. Solon (1999), Brock and Durlauf (2001) and Ginther et al. (2000), provide recent surveys of this literature. Significant coefficients on the group-level variables are taken to indicate the presence of social, peer group or "neighborhood" effects.

The identifying assumptions underlying much of the research were first formally analyzed by Manski (1993), in his now seminal "reflection problem" paper, which established two non-identification results. First, it is in general impossible to distinguish the effects of social interactions from those driven unobserved group characteristics (see Manski (1993), Corollary to Proposition 1). Second, it is impossible to distinguish between endogenous social effects, where individual behavior varies with mean reference group behavior, and exogenous social effects, where individual behavior varies with mean group composition (see Manski, 1993, Proposition 1).

In this paper, we provide an extra source of information that can help identify the linear-in-means model of social interactions with unobserved group effects. We focus on identification of endogenous social effects from unobserved group characteristics under the assumption that exogenous social effects are not present[1]. We reinterpret the linear-in-means model as a quasi-panel data model, where the "cross sectional dimension" equals the number of observed social groups and the "time series dimension" equals the number of sampled individuals within each group. Using our quasi-panel reinterpretation, it is straightforward to see that Manski's first non-identification result for the linear-in-means model is analogous to the inability of a standard "fixed effects" regression to identify coefficients on group-invariant regressors. Exploiting the idea of Hausman and Taylor (1981) developed for panel data models, we identify the between-group variation that contains information on the social multiplier (Manski, 1993; Becker and Murphy, 2000; Glaeser et al., 2002). Existence of instruments generating exogenous between-group variation plays the role of the extra information needed to overturn the result of Manski (1993). Despite their simplicity, our results do not seem to be ex-ante obvious given the idiosyncratic state of empirical work.

2. Identification by a quasi-panel approach

In this section we discuss the linear-in-means models of social interactions. We begin by considering the case where the researcher has access to a grouped cross-section of data, where the econometrician observes N non-overlapping social groups individuals ($g=1, \ldots, N$) and where M^g individuals ($i=1, \ldots, M^g$) are sampled in the gth group. The following linear-in-means model of social interactions generates the observed data,

$$y_{gi} = E_g\left[y_{gi}\right]\beta + E_g\left[\boldsymbol{x}'_{gi}\right]\gamma + \boldsymbol{x}'_{gi}\delta + \boldsymbol{r}'_{gi}\eta + u_{gi}$$

$$u_{gi} = \alpha_g + \varepsilon_{gi} \tag{1}$$

where all variables are measured in deviations from sample means. Here, $E_g[\cdot]$ denotes the mean for the gth group. Let y_{gi} equal the outcome/behavior of interest for the ith individual in the gth group.

[1] In contrast, Brock and Durlauf (2001) focused on identification of endogenous social effects from exogenous social effects under the assumption that there does not exist any unobserved group characteristic. Their identification strategy utilized nonlinearity of the model, which cannot be exploited in linear models.

Following the terminology introduced by Manski (1993), the coefficient on $E_g[y_{gi}]$—the average outcome within the gth group—determines the strength of endogenous social effects in explaining individual outcomes. In addition to y_{gi}, we observe x_{gi}, a vector of individual characteristics that also generate exogenous (contextual) social effects, and $\boldsymbol{r}_{gi}$, a vector of individual characteristics that operate at the individual level only. Finally α_g, which is not observed by the econometrician, captures the presence of correlated group effects.

We will focus on a simple version of Eq. (1), where there are no exogenous social effects ($\dim(\mathbf{x}_{gi})=J=0$):

$$y_{gi} = E_g\left[y_{gi}\right]\beta + \boldsymbol{r}'_{gi}\eta + \alpha_g + \varepsilon_{gi}. \tag{2}$$

It is useful to compute several transformations of the structural model (2). These transformations will form the basis of our identification and estimation strategy.

Taking group means of both sides of (2) and solving for $E_g[y_{gi}]$ yields the social equilibrium (c.f., Manski, 1993):

$$E_g\left[y_{gi}\right] = E_g\left[\boldsymbol{r}'_{gi}\right]\frac{\eta}{1-\beta} + \frac{\alpha_g}{1-\beta} \tag{3}$$

where we assume $E_g[\varepsilon_{gi}]=0$ without loss of generality. Throughout we also assume that every group is in social equilibrium, i.e., $E_g[y_{gi}]$ satisfies (3) for all g. Under this assumption, we can work with a reduced form version of the linear-in-means model[2]. Substituting the social equilibrium (3) into the structural model (2) results in

$$y_{gi} = \boldsymbol{r}'_{gi}\eta + E_g\left[\boldsymbol{r}'_{gi}\right]\frac{\beta\eta}{1-\beta} + \alpha_g^* + \varepsilon_{gi} \qquad g=1,\ldots,N, \quad i=1,\ldots,M^g \tag{4}$$

where $\alpha_g^* = \frac{\alpha_g}{1-\beta}$.

Suppose tentatively that we observe $E_g[\boldsymbol{r}_{gi}]$ for each group. We can then identify β and η if we can identify the coefficients on $E_g[\boldsymbol{r}_{gi}]$ and $\boldsymbol{r}_{gi}$. Regarding Eq. (4) as a quasi-panel model, it is clear that a simple fixed effects approach will not identify β because $E_g[\boldsymbol{r}_{gi}]$ plays a role analogous to a time-invariant regressor in a standard panel data model. The group-invariant social interactions vector, $E_g[\boldsymbol{r}_{gi}]$, will be eliminated by the within-group transformation, which forms the basis of the fixed effects approach. Hence viewing (4) as a quasi-panel data model provides an intuitive reinterpretation of Manski (1993) main non-identification result appropriate to this sampling structure. Fortunately, it is well-known that a generalization of the instrumental variables strategy of Hausman and Taylor (1981) can be used to identify the coefficient of a time-invariant regressor in a standard panel data model. Our identification strategy exploits this type of intuition.

In order to discuss identification and estimation it will prove useful to rewrite the reduced form, (4), in terms of observables only as

$$y_{gi} = \boldsymbol{r}'_{gi}\eta + \overline{\boldsymbol{r}}'_g\frac{\beta\eta}{1-\beta} + u_{gi} \tag{5}$$

[2] The social equilibrium assumption is strong, but it eliminates the intrinsic simultaneity problem that characterizes the structural model.

where

$$u_{gi}=\frac{\alpha_g}{1-\beta}-\left(\bar{\boldsymbol{r}}'_g-E_g\left[\boldsymbol{r}'_{gi}\right]\right)\frac{\beta\eta}{1-\beta}+\varepsilon_{gi}.$$

Note that replacing the unobserved $E_g[\boldsymbol{r}_{gi}]$ vector with its sample analogue, $\bar{\boldsymbol{r}}_g$, creates an error-in-variables problem.

We first discuss how η can be identified by the within-group variation. The within-group version of the reduced form is obtained by inspecting (4):

$$\tilde{y}_{gi}=\tilde{\boldsymbol{r}}'_{gi}\eta+\tilde{\varepsilon}_{gi}, \tag{6}$$

where for any vector $\boldsymbol{s}_{gi}$, $\tilde{\boldsymbol{s}}_{gi}\equiv\boldsymbol{s}_{gi}-\bar{\boldsymbol{s}}_g$. Note that η is identified as the limit of the OLS estimator of $\tilde{y}_{gi}$ on $\tilde{\boldsymbol{r}}_{gi}$ under a standard strict exogeneity condition on $\boldsymbol{r}_{gi}$: If $E\left[\varepsilon_{gi}|\boldsymbol{r}_{g1},\ldots,\boldsymbol{r}_{gM^g},E_g\left[\boldsymbol{r}_{gi}\right],\alpha_g\right]=0$, then we have

$$\eta=\left(E\left[\sum_{i=1}^{M^g}\tilde{\boldsymbol{r}}_{gi}\tilde{\boldsymbol{r}}'_{gi}\right]\right)^{-1}E\left[\sum_{i=1}^{M^g}\tilde{\boldsymbol{r}}_{gi}\tilde{y}_{gi}\right]=\operatorname{plim}\left[\frac{1}{N}\sum_{g=1}^{N}\left(\sum_{i=1}^{M^g}\tilde{\boldsymbol{r}}_{gi}\tilde{\boldsymbol{r}}'_{gi}\right)\right]^{-1}\left[\frac{1}{N}\sum_{g=1}^{N}\left(\sum_{i=1}^{M^g}\tilde{\boldsymbol{r}}_{gi}\tilde{y}_{gi}\right)\right] \tag{7}$$

We now discuss how $\frac{\eta}{1-\beta}$ can be identied by the between-group variation. From the reduced form (5) we note that $\bar{y}_g$ equals:

$$\bar{y}_g=\bar{\boldsymbol{r}}'_g\frac{\eta}{1-\beta}+\bar{u}_g \tag{8}$$

This is the between-group version of the reduced form linear-in-means model. We assume existence of an instrument q_g that satisfy the following standard conditions of instrument validity: $E[q_g\bar{u}_g]=0$,$E[q_gE_g[\boldsymbol{r}'_{gi}]]\neq 0$,$\dim(q_g)\geq\dim(\boldsymbol{r}_{gi})$. It then follows that

$$\frac{\eta}{1-\beta}=\operatorname{plim}\left\{\left(\frac{1}{N}\sum_{g=1}^{N}\bar{\boldsymbol{r}}_g q_g\right)\left(\frac{1}{N}\sum_{g=1}^{N}q_g q'_g\right)^{-1}\left(\frac{1}{N}\sum_{g=1}^{N}q_g\bar{\boldsymbol{r}}'_g\right)\right\}^{-1}$$
$$\times\left(\frac{1}{N}\sum_{g=1}^{N}\bar{\boldsymbol{r}}_g q_g\right)\left(\frac{1}{N}\sum_{g=1}^{N}q_g q'_g\right)^{-1}\left(\frac{1}{N}\sum_{g=1}^{N}q_g\bar{y}_g\right). \tag{9}$$

The limit of the within OLS estimator (7) identifies η, while $\frac{\eta}{1-\beta}$ can be identified as the limit of the between 2SLS estimator (9). These two sources of identification can then be combined to produce an estimator for β. Intuitively, our identification strategy exploits a basic prediction of the linear-in-means model to achieve identification: in the presence of social effects, the impact of covariate differences on the between-group variation in outcomes should be greater than the corresponding effect on within-group variation. This prediction accords well with the notion that social interactions generate “excessive variation” in outcomes. Put differently, the relationship between $\bar{y}_g$ and $\bar{\boldsymbol{r}}_g$ exhibits a “multiplier” effect, which is our source of identication for β (Lewis, 1963; Becker and Murphy, 2000; Glaeser et al., 2002). The success of this identification strategy crucially hinges on the availability on an exogenous source of between-group variation in $\bar{\boldsymbol{r}}_g$. Our instrument based identification strategy is one way of extracting such variation.

Thus far we have only explored identification of the linear-in-means model without exogenous social effects. When some exclusion restriction is available, we can in principle identify exogenous social effects as well. Suppose that

$$y_{gi} = \beta E_g\left[y_{gi}\right] + E_g\left[\boldsymbol{x}'_{gi}\right]\gamma + \boldsymbol{x}'_{gi}\delta + \boldsymbol{r}'_{gi}\eta + \alpha_g + \varepsilon_{gi}$$

Then we can identify the exogenous social effects γ and the endogenous social effects β from the reduced form

$$y_{gi} = \boldsymbol{x}'_{gi}\delta + \boldsymbol{r}'_{gi}\eta + E_g\left[\boldsymbol{x}'_{gi}\right]\frac{\gamma + \beta\delta}{1-\beta} + E_g\left[\boldsymbol{r}'_{gi}\right]\frac{\beta\eta}{1-\beta} + \frac{\alpha_g}{1-\beta} + \varepsilon_{gi}$$

which can be identified by methods discussed before. Note that we are assuming the presence of a vector of individual-level covariates, $\boldsymbol{r}_{gi}$, that operate at the individual-level only and hence do not generate contextual effects (Moffitt, 2001; Brock and Durlauf, 2001). Such exclusion restriction will probably be difficult to justify in practice, though.

3. Summary

This paper has presented new results on identification and estimation for the linear-in-means model of social interactions. Our main innovation is to illustrate that the linear-in-means model can be reinterpreted as a quasi-panel model. The advantages of this reinterpretation include transparent and intuitive identifying conditions and an associated GMM framework for estimation and inference.

Acknowledgments

We would like to thank William Brock, Gary Chamberlain, Jerry Hausman, Enrico Moretti and participants at UCLA Workshop in Economic Applications and the Harvard development and econometric lunches for many helpful comments. A very special thanks goes to Ariel Pakes, who appeared as a co-author on an initial version of this paper, and whose insight and discussions have proved invaluable. We are much indebted to his intellectual generosity. The second author gratefully acknowledges Financial support from NSF Grant SES-0313651.

References

Becker, Gary, Murphy, Kevin M., 2000. Social Economics: Market Behavior in a Social Environment. Harvard University Press, Cambridge, MA.

Brock, William A., Durlauf, Steven N., 2001. Interactions-based models. In: Heckman, J., Leamer, E. (Eds.), Handbook of Econometrics, vol. 5. North-Holland, Amsterdam, pp. 3297–3380.

Ginther, Donna, Haveman, Robert, Wolfe, Barbara, 2000. Neighborhood attributes as determinants of children's outcomes: how robust are the relationships. Journal of Human Resources 35 (4), 603–642.

Glaeser, Edward L., Sacerdote Bruce and Jose A. Scheinkman (2002). "The Social Multiplier", NBER Working Paper No. 9153.
Hausman, Jerry A., Taylor, William E., 1981. Panel data and unobservable individual effects. Econometrica 49 (6), 1377–1398 (Nov.).
Lewis, Gregory H., 1963. Unionism and Relative Wages in the United States: An Empirical Inquiry. Chicago University Press, Chicago.
Manski, Charles F., 1993. Identification of endogenous social effects: the reflection problem. Review of Economic Studies 60 (3), 531–542.
Moffitt, Robert A., 2001. Policy interventions, low-level equilibria and social interactions. In: Durlauf, S., Young, P. (Eds.), Social Dynamics. MIT Press, Cambridge, MA.
Solon, Gary, 1999. Intergenerational mobility in the labor market. In: Ashenfelter, O., Card, D. (Eds.), Handbook of Labor Economics vol. 3. North-Holland, Amsterdam, pp. 1761–1800.