Social Interactions: Identification

Steven N. Durlauf

University of Chicago

SSSI 2017

In these notes, I discuss identification problems in the uncovering of empirical evidence on social interactions.

The workhorse of empirical research on social interactions is the linear in means model.

The version of this model which has been the focus of most identification analysis is due to Manski (1993).

The model assume that individuals are arrayed across nonoverlapping groups g and make choices ω_i following

$$\omega_i = k + cx_i + dy_g + Jm_{i,g}^e + \varepsilon_i \qquad (1)$$

where x_i denotes a $r-$ length vector of individual characteristics, y_g denotes an $s-$ length vector of group characteristics, sometimes cannled contextual effects, and. $m_{i,g}^e$ denotes the expected average behavior of others in the group, i.e.

$$m_{i,g}^e = \frac{1}{n_g}\sum_{j \in g} E\left(\omega_j \middle| x, y_g\right) \qquad (2)$$

Note that (2) ignores the effect of an individual's ε_i on his beliefs about the average choice. As such, (2) is not strictly rational. As will be discussed, the linear in means model as described here is an approximation of a microfounded model.

Claims about social interactions are, from the econometric perspective, equivalent to statements about the values of d and J.

The statement that social interactions matter is equivalent to the statement that at least some element of the union of the parameters in d and J is nonzero. The statement that contextual social interactions are present means that at least one element of d is nonzero.

The statement that endogenous social interactions matter means that J is nonzero.

In Manski's original formulation, $y_g = \bar{x}_g$, where $\bar{x}_g = \frac{1}{n_g}\sum_{i \in g} x_i$ denotes the average across i of x_i within a given g, which explains the model's name. Regardless of whether they are equal, I assume that both y_g and $\bar{x}_g$ are observable to individuals and discuss how to relax this below.

Assumptions on Errors

First the expected value of ε_i is 0, conditional on the information set $\left(x_i, \overline{x}_g, y_g, i \in g\right)$[1]

$$\forall i,g \; E\left(\varepsilon_i \middle| x_i, \overline{x}_g, y_g, i \in g\right) = 0 \qquad (3)$$

Second

$$\text{For each } i, j, g, h \text{ such that } i \neq j \text{ or } g \neq h$$
$$cov\left(\varepsilon_i \varepsilon_j \middle| x_i, \overline{x}_g, y_g, i \in g, x_j, \overline{x}_h, y_h, j \in h\right) = 0 \qquad (4)$$

[1]The conditioning argument means that one is conditioning on the fact that i is a member of group g.

Eq. (4) eliminates conditional cova $i \in g$ riation between the errors. The inclusion of the group memberships, e.g. $i \in g$ rules out some relationship between the identity of the group and model errors, thereby allowing us to treat groups as exchangeable.

Eqs. (1)-(4) imply that agents in a group have common expectations about the expected average choice. Under rational expectations,

$$m_g^e = m_g \equiv \frac{k + c\bar{x}_g + dy_g}{1 - J} \qquad (5)$$

The equation says that the individuals' expectation of average behavior in the group equals the average behavior of the group, and this in turn depends linearly on the average of the individual determinants of behavior, $\bar{x}_g$, and the contextual interactions that the group members experience in common. The condition $J < 1$ is required for equation to make sense.

Reduced form

Substitution of (5) into (1) eliminates m_g and so provides a reduced form version of the linear in means model in that the individual outcomes are determined entirely by observables and the individual-specific error:

$$\omega_{i,g} = \frac{k}{1-J} + cx_i + \frac{J}{1-J} c\bar{x}_g + \frac{d}{1-J} y_g + \varepsilon_i . \qquad (6)$$

This reduced form corresponds to the bulk of the empirical literature which has focused on the regression

$$\omega_{i,g} = \pi_0 + \pi_1 x_i + \pi_2 y_g + \varepsilon_i . \qquad (7)$$

where the parameters π_0, π_1, π_2 are taken as the objects of interest

A comparison of (7) with (6) indicates how findings in the empirical literature that end with the reporting of π_0, π_1, π_2 inadequately address the task of fully characterizing the social interactions that are present in the data.

Instrumental variables and the reflection problem

As follows from (6), if $\bar{\omega}_g$ is projected against the union of elements of $\bar{x}_g$ and y_g, this produces the population mean m_g. Hence, we can proceed as if m_g is observable.

Put differently, our identification arguments rely on the analogy principle which means that one works with population moments to construct identification arguments.

Since y_g appears in (1) it will not facilitate identification. As we shall see, identification via instrumental variables is determined by the informational content of $\bar{x}_g$ relative to y_g.

As first recognized by Manski (1993), identification can fail for the linear in means model when one focuses on the mapping from reduced form regression parameters to the structural parameters. Manski's assumes $y_g = \bar{x}_g$. In this case, eq. (5) reduces to

$$m_g = \frac{k + (c+d)y_g}{1-J} \tag{8}$$

The regressor m_g in (1) is linearly dependent on the other regressors, i.e. the constant and y_g. This linear dependence is the reason that identification fails: the comovements of m_g and y_g are such that one cannot disentangle their respective influences on individuals. Manski (1993) named this failure the reflection problem. Metaphorically, if one observes that $\omega_{i,g}$ is correlated with the expected average behavior in a neighborhood, (8) indicates it may be possible that this correlation is due to the fact that m_g may simply reflect the role of y_g in influencing individuals.

As such, the reflection problem is a variant of the classical failure of identification in a simultaneous equations models.

Conditions for Identification

Under what conditions is this model identified?

A necessary condition is that Manski's assumption that $y_g = \bar{x}_g$ is relaxed. This will allow for the possibility m_g is not linearly dependent on the constant and y_g. The reason for this is the presence of the term $\frac{c\bar{x}_g}{1-J}$ in eq. (6). This term can break the reflection problem as m_g may not be linearly dependent on the other regressors in (6). This immediately leads to the argument in Brock and Durlauf (2001b) that a necessary condition for identification in the linear in means model, is that there exists at least one element of x_i whose group level average is not an element of y_g, while Durlauf and Tanaka (2008) provide a sufficient set of conditions.

Consider the projections $proj\left(\bar{\omega}_g \middle| 1, y_g, \bar{x}_g\right)$ and $proj\left(\bar{\omega}_g \middle| 1, y_g\right)$, where 1 is simply a random variable with mean 1 and variance 0, corresponding to the constant term. The first projection provides an optimal linear forecast (in the variance minimizing sense) of $\bar{\omega}_g$, conditioning on the random variables defined by 1 and the elements of y_g and $\bar{x}_g$, whereas the second projection provides the optimal linear forecast when only 1 and the elements of y_g are used. The difference between the two projections thus measures the additional contribution to predicting $\bar{\omega}_g$ beyond what can be achieved using $\bar{x}_g$ in addition to 1 and y_g. When this marginal contribution is nonzero, then it is possible to to identify the structural parameters in (1) using instrumental variables for $\bar{\omega}_g$ or equivalently for m_g. Formally,

Theorem 1. Identification of the linear in means model of social interactions.

The parameters (k,c,J,d) are identified if and only if $proj\left(\bar{\omega}_g \middle| 1, y_g, \bar{x}_g\right) - proj\left(\bar{\omega}_g \middle| 1, y_g\right) \neq 0$

The intuition for the theorem is simple. Identification requires that one can project $\bar{\omega}_g$ (equivalently) onto a space of variables such that the projection is not collinear with the other regressors in the model.[2]

---

[2]The conditions of the theorem do not preclude a functional dependence of X_i on y_g, which, combined with the uniqueness of m_g, means that the nonparametric analog to the model is not identified, following Manski (1993, Proposition 3). This observation builds on discussion in Manski (1993, p. 539).

Theorem 1 was derived under the assumption that $\bar{x}_g$ and y_g are known to the individual decisionmakers at the time that their choices are made. This assumption is a strong one and further may appear to be inconsistent with our assumption that $\bar{\omega}_g$ is unobservable to them. This latter concern is not tenable: in a context such as residential neighborhoods, it is possible for a contextual effect such as average income to be observable whereas the school effort levels of children in the neighborhood are not. However, it is important to understand the implications of relaxing our baseline informational assumptions on identification. This is the contribution of Graham and Hahn (2005).

The models they study can be subsumed as variants of a modified version of (1)

$$\omega_{i,g} = k + cx_i + dE\left(y_g \middle| F\right) + Jm_g + \varepsilon_i \qquad (9)$$

where individuals are assumed to possess a common information set F. As such, it is clear that the conditions for identification in Theorem 1 are easily generalized. One simply needs a set of additional instruments q_g such that the elements of q_g can jointly instrument $E\left(y_g\right)$ and m_g. As they observe, the variables q_g constitute exclusion restrictions and so require prior information on the part of the analyst. For their context, y_g is a strict subset of $\bar{x}_g$, so it is difficult to justify the observability of those elements of $\bar{x}_g$ that do not appear in y_g when the others are by assumption not observable.

In our view, the appropriate route to uncovering valid instruments q_g, under the Graham and Hahn information assumptions, most likely requires the development of an auxiliary model of x_i and hence $\bar{x}_g$.

In other words, Graham and Hahn's concerns reflect the incompleteness of (9) in the sense that the individual characteristics are not themselves modeled. Hence, we interpret their argument as one that calls for the embedding of outcomes such as (9) in a richer system, possibly one including dynamics, which describes how individual characteristics are determined. We fully agree with Graham and Hahn that in isolation, finding valid instruments for (9) is difficult, but would argue that this difficulty reflects the limitations of studying $\omega_{i,g}$ in isolation rather than as one of a set of equilibrium outcomes.

Variations of the linear in means model

We now evaluate the reflection problem for some econometric models that differ from (1) in various ways that are common in empirical work.

Once one considers econometric structures outside the linear cross-section framework, the reflection problem may not arise, even if there is a one-to-one correspondence between individual and contextual interactions.

We consider three alternative structures.

Partial linear-in-means models

The linear structure in (1) is theoretically justified under strong function form assumptions for utility, which leads to the question of whether relaxation of the linearity assumption affects identification. One such relaxation is studied in Brock and Durlauf (2001) and involves a particular nonlinear generalization under rational expectations:

$$\omega_{i,g} = k + cx_i + dy_g + J\mu\left(m_g\right) + \varepsilon_i \qquad (10)$$

This type of structure is known as a partial linear model.

Brock and Durlauf establish that the parameters of this model are identified for those elements of the space of twice differentiable functions, for known $\mu\left(m_g\right)$, so long as $\frac{\partial^2 \mu\left(m_g\right)}{\partial m_g^2} \neq 0$, outside of nongeneric cases.

The intuition is straightforward; the reflection problem requires linear dependence between group outcomes and certain group-level aggregates, which is ruled out by the nonlinearity in (10). We should note that there does not exist any identification results, as far as we know, if the functional form for $\mu\left(m_g\right)$ is unknown, so in this sense the identification of (10) does not exploit results from the semiparametric literature on partial linear models.

The finding that partial linear variants of (1) do not suffer from the reflection problem is not a surprise from the perspective of the simultaneous equations literature.

McManus (1992), in what appears to be an underappreciated paper, illustrates how for a broad class of parametric nonlinear simultaneous equations models, subsets of nonidentified models are nongeneric. For example, McManus (1992, pg. 8) shows in his pedagogical example that "…First the set of δ values which correspond to identified (non identified) models forms an open and dense (nowhere dense) subset of the real line…" He develops a general argument which formalizes this basic idea. Brock and Durlauf (2001, p. 3371) adapt McManus's argument to show that "…the local nonidentification of the linear-in-means model can be perturbed away by a C^2-small change." See Brock and Durlauf (2001) for the details of this extension to social interactions models.

Dynamic linear models

Similarly, dynamic analogs of the linear in means model may not exhibit the reflection problem. Brock and Durlauf (2001) illustrate this with the dynamic social interactions model

$$\omega_{i,g,t} = k + cx_{i,t} + dy_{g,t} + \beta m_{g,t-1} + \varepsilon_{i,t}. \qquad (11)$$

This model avoids linear dependence between the contextual and endogenous variables since

$$m_{g,t} = \frac{k + c\bar{x}_{g,t} + dy_{g,t}}{1 - \beta L} \qquad (12)$$

where L is a lag operator. Eq. (12) implies that $m_{g,t}$ depends on the entire history of $\bar{x}_{g,t}$ and $y_{g,t}$. This model is essentially backwards looking and is driven by the idea that current behaviors are directly affected by past beliefs.

Similar results hold for forward looking models. An example of a model in this class is

$$\omega_{i,g,t} = k + cx_{i,t} + dy_{g,t} + \beta m_{g,t+1} + \varepsilon_{i,t} \qquad (13)$$

This model is equivalent to the workhorse geometric discount model in rational expectations (Hansen and Sargent (1980)). The equilibrium average choice level for a group equals, following Hansen and Sargent,

$$m_{g,t} = \frac{k}{1-\beta} + \sum_{s=0}^{\infty} \beta^s E_t\left(c\bar{x}_{g,t+s} + dy_{g,t+s}\right) \qquad (14)$$

It is immediate that regressors are linearly independent so long as $\bar{x}_{g,t}$ and $y_{g,t}$ are not both random walks.

Identification of this class of dynamic models was originally studied in Wallis (1980).

Hierarchical models

In fields such as sociology, social interactions are typically explored using hierarchical models, i.e. models in which contextual interactions alter the coefficients that link individual characteristics to outcomes. See Bryk and Raudenbush (2001) for a full description of the method. T

The reason for this appears to be a different conceptualization of the meaning of social interactions in economics in comparison to other social sciences.

Hierarchical models appear, in our reading, to be motivated by a view of social groups as defining ecologies in which decisions are made and matter because different social backgrounds induce different mappings from the individual determinants of these behaviors and choices, cf. Raudenbush and Sampson (1999).

Economics, in contrast, regards the elements that coTmprise endogenous and contextual social interactions as directly affecting the preferences, constraints, and beliefs of agents and so treats them as additional determinants to individual specific characteristics, x_i.

There do not exist formal arguments for favoring one approach versus another at an abstract level.

At the same time the additivity assumption in both approaches are ad hoc from the perspective of economic theory, even if the assumption is ubiquitous in empirical practice.

For hierarchical models, there has been no virtually attention to the reflection problem except Blume and Durlauf (2005). Here we modify the Blume and Durlauf analysis and consider a formulation that closely follows the conceptual logic of hierarchical models in that social interactions are entirely subsumed in the interactions on parameters. Formally, this means that individual outcomes obey

$$\omega_{i,g} = k_g + c_g x_i + \varepsilon_i \qquad (15)$$

With individual- and group-specific components obeying

$$k_g = k + dy_g + Jm_g \qquad (16)$$

and

$$c_g = c + y_g'\Psi + m_g\psi \qquad (17)$$

Ψ is a matrix and ψ is a vector. I omit any random terms in (17) and (18) for simplicity, although hierarchical models typically include them.

This formulation assumes that the endogenous effect directly affects the individual level coefficients and so differs from the Blume and Durlauf example. Imposing rational expectations, the hierarchical model is equivalent to the linear model

$$\omega_{i,g} = k + cx_i + dy_g + Jm_g + y_g'\Psi x_i + m_g\psi x_i + \varepsilon_i \quad (18)$$

Hence, the difference between the linear model used in economics and the hierarchical structure is the addition of the terms $y'_g \Psi x_i$ and $m_g \psi x_i$ by the hierarchical model to the original linear in means model.

Thus the hierarchical model does nothing deeper than add the cross products of variables in order to allow for nonlinearity. As such, the approach is far behind the econometrics literature on semiparametric methods which allows for much deeper forms of nonlinearity.

On the other hand, the use of cross products of variables is common in empirical economics.

Can this model exhibit the reflection problem? The self-consistent solution expected choice level for the hierarchical model is

$$m_g = \frac{k + c\bar{x}_g + dy_g + y_g'\Psi\bar{x}_g}{1 - J - \psi\bar{x}_g} \qquad (19)$$

Recall that the reflection problem occurred when $y_g = \bar{x}_g$. If we impose this condition in the hierarchical model, (20) becomes

$$m_g = \frac{k + (c+d)y_g + y_g'\Psi y_g}{1 - J - \psi y_g} \qquad (20)$$

Equation (20) makes clear that the relationship between m_g and the other regressors in the hierarchical model is nonlinear.

The presence of $y_g'\Psi y_g$ in the numerator and $-\psi y_g$ in the denominator ensures that linear dependence will not hold, except for hairline cases, so long as there is sufficient variation in x_i and y_g.

Hierarchical models thus exhibit different identification properties from linear in means models because their structure renders the endogenous effect m_g a nonlinear function of the contextual interactions y_g (and also a nonlinear function of $\bar{x}_g$ if this variable is distinct from y_g).

U

nobserved group effects

One of the major limits to identification of social interactions is the presence of unobserved group-level heterogeneity. To introduce this issue, we modify (1) to

$$\omega_{i,g} = k + cx_i + dy_g + Jm_g + \alpha_g + \varepsilon_i \qquad (21)$$

The associated reduced form for (22) is

$$\omega_{i,g} = \frac{k}{1-J} + cx_i + \frac{J}{1-J}c\bar{x}_g + \frac{d}{1-J}y_g + \frac{1}{1-J}\alpha_g + \varepsilon_i. \quad (22)$$

Instrumental variables

One approach to dealing with unobserved group level heterogeneity is the use of instrumental variables. This approach is generally difficult to justify in addressing unobserved group characteristics for both the linear in means and other models.

The reason for the difficulty is that α_g is itself undertheorized, in other words, this term captures aspects of a group that affect outcomes which the model does not explicitly describe.

Beyond this, valid instrumental variables require the property that they have been excluded from the initial behavioral equation as either individual or contextual determinants of outcomes.

It is hard to see how, in typical socioeconomic contexts, such instruments may be found, since the instruments must be known on a priori grounds to be uncorrelated with both the undertheorized α_g and ε_i.

Social interactions models are typically what Brock and Durlauf (2001) have termed openended, which means that their theoretical structure does not naturally identify variables to exclude from the equations that describe behavior.

In other words, social interactions theories are openended because the presence of a given type of social interaction does not logically preclude the empirical relevance of other theories; the econometric analog of this is that social economics models do not provide a logical basis for choosing instruments.

This is quite different from rational expectations models, for example, whose logic often allows one to express linear combinations of variables as forecast errors, which must logically be orthogonal to an agent's information set; in macroeconomics a key example of this is the Euler equation in a stochastic optimization model.

Some uses of instrumental variables fall under the rubric of quasi-natural experiments. A recent example is Cipollone and Rosolia (2007) which we describe in some detail as it illustrates the strengths and weaknesses of quasi-experimental data as a source for evidence on social interactions.

Their analysis examines the effects of changes in male high school graduation rates on female high school graduation rates using a change in Italy's compulsory military service laws which exempted male students in schools located in areas damaged by a 1980 earthquake.

Cipollone and Rosolia compare two groups of schools.

The first group of schools are located in towns that experienced relatively little earthquake damage (based on official assessments) and yet were included in the draft exemption.

The second group of schools were located in towns that were near the towns whose schools comprise the first group; the authors argue that these towns suffered similar damage so that their failure to receive an exemption was arbitrary. Cipollone and Rosolia find statistically significant higher graduation rates for females in the high schools subject to the exemption when compared to females in the comparable high schools that were not subject to exemption.

One limitation of this type of calculation is that it is difficult to interpret in terms of social mechanisms, an issue recognized by the authors.

Regardless of this, the finding itself may be problematic. Compulsory military service was previously subject to exemptions for high school graduation. Thus, the general exemption changed the composition of males in a school in particular ways. The problem is that the general exception affected the attendance of males whose unobservable characteristics made their graduation behavior especially sensitive to the policy change relative to the previous regime.

Suppose that there is assortative matching on these unobservable characteristics in the formation of romantic relationships.

One can reasonably imagine that an associated increase in graduation for females occurs because of the preservation of romantic relationships that would have been severed by school (and community) withdrawal for military service. The message of this possibility is that the translation of what amount to partial correlations on the behavior of one group with the behaviors of another group into causal claims about social interactions that can answer policy relevant questions requires careful consideration of counterfactuals and the nature of unobservable individual-specific heterogeneity.

Panel data

A second standard strategy for dealing with unobserved group interactions involves the use of panel data to difference the interactions out. This amounts to working with

$$\omega_{i,g,t} - \omega_{i,g,t-1} = c\left(x_{i,t} - x_{i,t-1}\right) + d\left(y_{g,t} - y_{g,t-1}\right) + J\left(m_{g,t} - m_{g,t-1}\right) + \varepsilon_{i,t} - \varepsilon_{i,t-1} \qquad (23)$$

Recall that our identification Theorem 1 depended on the relationship between $\bar{x}_g$, y_g and m_g. Theorem 1 immediately can be applied if one considers the requirements of the Theorem as they apply to $\bar{x}_{g,t} - \bar{x}_{g,t-1}$, $y_{g,t} - y_{g,t-1}$ and $m_{g,t} = m_{g,t-1}$.

So long as there is temporal variation in $\bar{x}_{g,t}$ and $y_{g,t}$ i.e. the first differences are not zero, then the conditions for identification will be the same as in the original linear model without α_g. Note that variation in $\bar{x}_{g,t}$ and/or $y_{g,t}$ will induce variation in $m_{g,t}$ over time. An early example of this strategy is Hoxby (2000) who focuses on variation in the percentage of a student´s own ethnic group in a classroom.

Self-selection

It is natural for many social contexts to expect individuals to self-select into groups. This is most obvious for the case of residential neighborhoods; models such as Bénabou (1993,1996), Durlauf (1996a,b) and Hoff and Sen (2005), for example, all link social interactions to neighborhood choice. In terms of estimation, self-selection generally means that orthogonality of regressors and errors is violated.

Self-selection has typically been addressed using instrumental variables methods.

The use of instrumental variables as a solution to self-selection suffers, in our view, from the problem of theory openendedness as was discussed in the context of unobserved group effects.

However, unlike the case of unobservable group interactions, self-selection involves a specific behavior on the part of the agents under study which can provide additional insight into instrument validity.

For example, Evans, Oates and Schwab (1992) focus on estimating the effect of the percentage of students in a school who are disadvantaged on high school dropout and teen fertility rates. The measure of school level socioeconomic disadvantage is instrumented with metropolitan area levels of unemployment, college completion and poverty rates and median income. The instruments are justified on the grounds that while families may choose schools within a metropolitan area, they are unlikely to choose metropolitan areas because of schools. This may be correct as far as it goes, but the relevant question for instrument validity is whether the instruments are uncorrelated with ε_i. One obvious reason why this is true is that drop out and pregnancy decisions will be related to labor market opportunities, which by the logic of Evans, Oates and Schwab's choice of the instruments would be defined at the metropolitan and not the school level. Durlauf (2004), on the other hand, suggests reasons why the instruments may not be valid.

In our view, the preferred approach to dealing with self-selection is to treat group choice and behavior within a group as a set of joint outcomes, and conduct empirical analysis from the perspective of both behaviors. Unlike the instrumental variables approach, this has interesting implications for identification, at least for the linear model; Brock and Durlauf (2001), first recognized this possibility and studied the case of self-selection between two groups; Brock and Durlauf (2002,2006) and Ioannides and Zabel (2008) extended this analysis to an arbitrary finite number of groups. At an intuitive level, this is not surprising. Self-selection represents a behavior on the part of an agent and so should contain information about his preferences, which will depend on the social interactions that occur in groups over which he is choosing. Unlike the instrumental variable approach, modeling self-selection exploits this information rather than treats it as a nuisance.

Following Heckman's original (1979) reasoning, one can think of individuals choosing between groups $g = 1,...,G$ based on an overall individual-specific quality measure for each group:

$$I^*_{i,g} = \gamma_1 x_i + \gamma_2 y_g + \gamma_3 z_{i,g} + \nu_{i,g} \qquad (24)$$

where $z_{i,g}$ denotes those observable characteristics that influence i's evaluation of group g but are not direct determinants of ω_i and $\nu_{i,g}$ denotes an unobservable individual-specific group quality term. Individual i chooses the group with the highest $I^*_{i,g}$. We assume that $\forall i,g$, $E\left(\varepsilon_i \middle| x_i, y_g, z_{i,g}\right) = 0$ and $E\left(\nu_{i,g} \middle| x_i, y_g, z_{i,g}\right) = 0$.

From this vantage point, self selection matters for identification because

$$E\left(\varepsilon_i \middle| x_i, \bar{x}_1, y_1, z_{i,1}, \ldots, \bar{x}_G, y_G, z_{i,G}, i \in g\right) \neq 0 \qquad (25)$$

Notice that eq. (25) includes the characteristics of all groups; this conditioning reflects the fact that the choice of group g depends on characteristics of the groups that were not chosen in addition to the characteristics of the group that was chosen.

Eq. (25) suggests that the linear in means model, under self-selection, should be written as

$$\omega_{i,g} = cx_i + dy_g + Jm_g + E\left(\varepsilon_i \middle| x_i, \bar{x}_1, y_1, z_{i,1}, \ldots., \bar{x}_G, y_G, z_{i,G}, i \in g\right) + \xi_i \quad (26)$$

where by construction $E\left(\xi_i \middle| x_i, \bar{x}_1, y_1, z_{i,1}, \ldots., \bar{x}_G, y_G, z_{i,G}, i \in g\right) = 0$. Notice that the conditioning in (25) includes the characteristics of all groups in the choice set; this is natural since the characteristics of those groups not chosen are informative about the errors.

This captures Heckman's (1979) insight that in the presence of self-selection on unobservables, the regression residual ε_i no longer has a conditional mean of zero, yet (25) can be consistently estimated using ordinary least squares if one adds a term to the original linear in means model ithat is proportional to the conditional expectation $E\left(\varepsilon_i \middle| x_i, \bar{x}_1, y_1, z_{i,1}, ..., \bar{x}_G, y_G, z_{i,G}, i \in g\right)$, prior to estimation. Denote this estimate as

$$\widehat{\kappa E\left(\varepsilon_i \middle| x_i, \bar{x}_1, y_1, z_{i,1}, ..., \bar{x}_G, y_G, z_{i,G}, i \in g\right)} \qquad (27)$$

Heckman's fundamental insight was that one can construct such a term by explicitly modeling the choice of group.

From this perspective, controlling for self-selection amounts to estimating

$$\omega_{i,g} = cx_i + dy_g + Jm_g + \rho\kappa \widehat{E\left(\varepsilon_i \middle| x_i, \bar{x}_1, y_1, z_{i,1}, \ldots, \bar{x}_G, y_G, z_{i,G}, i \in g\right)} + \xi_i \qquad (28)$$

Thus, accounting for self-selection necessitates considering identification for this regression.

The property of interest for the identification of social interactions is that the addition of the term can help facilitate identification.

To see this, consider two possible reasons why agents choose particular groups.

First, agents may choose groups on the basis of the expected average behaviors that occur. For example a family chooses a neighborhood based on its expectation of the average test score among students in the school their child will attend.

In the extreme case where this is the only neighborhood factor that matters to families, the conditional expectation associated with the selection correction will be a function of the agent's characteristics and the expected outcomes in each of the neighborhoods, i.e.

$$E\left(\varepsilon_i \middle| x_i, \bar{x}_1, y_1, z_{i,1}, \ldots, \bar{x}_G, y_G, z_{i,G}, i \in g\right) = \varpi\left(x_i, m_1, \ldots, m_G\right) \quad (29)$$

By the same logic that rendered the partial linear model identified, our equation is also identified as m_g cannot, outside of nongeneric cases be linearly dependent on a constant term and y_g.

Second, parents may choose neighborhoods based on the mean incomes of families or some measure of the distribution of occupations among neighborhood adults. This can be justified on role model grounds. If neighborhoods are evaluated according to their contextual variables, then (29) functions as an additional individual-specific regressor whose group level average does not appear in the model without self-selection. Hence, following the argument about identification in linear in means models that was developed earlier, the presence of a regressor with a nonzero coefficient can allow for identification to occur. This route to identification has been successfully used in Ioannides and Zabel (2008) to identify social interactions in housing.

Binary Choice Models

Suppose that agents choose $\omega_i \in \{-1,1\}$. Expected payoffs follow

$$EV(\omega_i) = k\omega_i + cX_i\omega_i + dY_g\omega_i + Jm_g\omega_i + \varepsilon_i(\omega_i)$$

where

$$\mu\left(\varepsilon_i(-\omega_i) - \varepsilon_i(\omega_i) \le z\right) = \frac{1}{1+\exp(-z)}$$

It is immediate that

$$\Pr\left(\omega_i = 1 \middle| X_i, Y_g, g\right) = \frac{\exp\left(k + cx_i + dy_g + Jm_g\right)}{\exp\left(k + cx_i + dy_g + Jm_g\right) + \exp\left(k + cx_i + dy_g + Jm_g\right)} \quad (30)$$

where

$$m_g = \tanh\left(k + cx_i + dy_g + Jm_g\right)$$

Recall $\tanh(x) = \dfrac{\exp(x) - \exp(-x)}{\exp(x) + \exp(-x)}$.

Identification

1. The parameters of the binary choice model are identified. True for multinomial choice analogs.

2. If the error difference obeys any absolutely continuous distribution, the distribution function is identified, as are parameters, so long as X_i, Y_g have unbounded support.

Intuition: within a group, distribution identified following arguments by Manski. Across groups, with sufficient variation, regressors cannot be linearly dependent.

Fixed Effects

In binary choice model, suppose that payoff obey

$$EV(\omega_i) = k\omega_i + cx_i\omega_i + dy_g\omega_i + Jm_g\omega_i + \alpha_g\omega_i + \varepsilon_i(\omega_i)$$

If α_g's are fixed effects, then identification fails without additional assumptions

Partial Identification Under Shape Restrictions on Density of α_g

Unlike the linear model, the binary choice model admists multiple equilibria.

This gives a route to a form of partial identification. Suppose that there are two groups, g and g' such that $m_g < m_{g'}$ yet

$$\tanh\left(k + cx_i + dy_g\right) > \tanh\left(k + cx_i + dy_{g'}\right)$$

How can this be?

Two routes:

1. $\alpha_g < \alpha_{g'}$

Or

2. g is coordinated at a low average equilibrium compared to g'

Pattern Reversals

The possibility of multiple equilibria suggests a form of partial identification.

Perhaps data can reveal that J is large enough for multiple equilibria.

This can be done under “mild” assumptions on unobservables

Theorem: Stochastic Dominance

Let $F_{\alpha_g|Y_g}$ denote distribution function of α_g Assume that $F_{\alpha_g|Y_g}$ first order stochastically dominates $F_{\alpha_{g'}|Y_{g'}}$ if $y_g > y_{g'}$.

Then if $y_g > y_{g'}$ and , then $J > 0$ and large enough to produce multiple equilibria.

$m_g < m_{g'}$

Theorem 2: Unimodality

Suppose that dF_{α_g} is unimodal. If $J = 0$, then there exists a vector π such that $dF_{\pi y_g | m_g}$ is unimodal.

Heuristics of proof: set $\pi = d$. Impact of contextual effects is monotonic under smoothness of expected value of group when equilibrium is unique.

Comment: means are not unimodal under multiple equilibria. Error in literature.