Identification of Binary Choice Models with Social Interactions

William A. Brock and Steven N. Durlauf*

January 30, 2004

*Corresponding author. We thank the National Science Foundation, John D. and Catherine T. MacArthur Foundation, Vilas Trust, and University of Wisconsin Graduate School for financial support. Ritesh Banerjee, Ethan Cohen-Cole, Giacomo Rondina, and Giulio Zanella have provided outstanding research assistance.

Identification of Binary Choice Models with Social Interactions

Abstract

This paper provides a set of results on the econometric identifiability of binary choice models with social interactions. Social interactions models have achieved recent prominence as economic analyses have attempted to incorporate the social effects of group memberships in understanding individual decisionmaking. Our analysis moves beyond parametric identification results that have been obtained in the literature to consider the identifiability of model parameters when the distribution of random payoff terms is unknown. Further, we consider how identification is affected by the presence of unobservable payoff terms of various types as well as identification in the presence of certain forms of endogenous group membership. Our results suggest that identification may be achieved under assumptions that in certain contexts may be plausible.

William A. Brock
Department of Economics
University of Wisconsin
1180 Observatory Drive
Madison, WI 53706-1393
wbrock@ssc.wisc.edu

Steven N. Durlauf
Department of Economics
University of Wisconsin
1180 Observatory Drive
Madison, WI 53706-1393
sdurlauf@ssc.wisc.edu

1. Introduction

This paper analyzes the identifiability of social interactions for binary choice models. In social interactions models, individual decisions are assumed to depend on various characteristics of the groups of which a given agent is a member. Social interactions have become an increasingly influential component of economic reasoning; Brock and Durlauf (2001b) and Manski (2000) survey a range of contexts in which social interactions have been argued to explain individual and aggregate outcomes. From the perspective of economic theorizing, the study of social interactions is important as it integrates substantive sociological ideas with formal economic reasoning. Early applications of social interactions to substantive theoretical problems include analyses of patterns of residential segregation (Schelling (1971)) and racial inequality (Loury (1977)); recent contributions include volatility in financial markets, (Brock (1993)), cross-city variation in crime (Glaeser, Sacerdote, and Scheinkman (1996)), and welfare dependence (Nechyba (2001)). Social interactions models may also be understood as exploring the consequences for individuals of their location in social space (cf. Akerlof (1997)); as such they are a complement to various spatial approaches to modeling; see Anselin (2003) for a discussion of spatial approaches to externalities.

While empirical work has overall lagged behind theoretical analyses of social interactions, a growing number of empirical studies have explored aspects of such [1]dependence.[2] Examples include Crane (1991) and South and Baumer (2000) who explore how neighborhood parental characteristics affect teenage fertility, Bertrand, Luttmer, and Mullainathan (2000) who analyze the role of aggregate ethnic group participation on individual welfare participation, Aizer and Currie (2002), who show how the use of publicly funded prenatal care is linked to social networks, Ioannides and Zabel (2003a,2003b) who identify social determinants of housing demand, and Sirakaya (2002) who evaluates the role of social interactions in criminal behavior.

[1]Authors such as Conley and Topa (2002), Glaeser, Sacerdote, and Scheinkman (1996) and Topa (2001) consider identification of social interactions using data that are aggregated at a group level.

[2] Empirical work on social interaction effects is surveyed in Durlauf (2003).

The purpose of this paper is to develop an analysis of the identification problem for social interactions in binary choice models using individual level data. Binary choice environments represent an important leading case for social interactions. One reason that the binary choice context is interesting is that many of the behaviors in which social interactions have been posited to matter, e,g, nonmartial fertility, commission of a crime, use of cigarettes, are binary in nature. In addition, binary choice models of social interactions possess a number of interesting theoretical properties, such as multiple equilibria and phase transition (the potential for qualitative changes in the properties of the model to change with small changes in model parameters) that are not present in linear-in means models.

Relative to previous work, we establish that the identification of social interactions in a binary choice context may be achieved under weaker assumptions than have appeared in the literature. For example, we show that identification does not require prior knowledge of the parametric specification of the distribution function for random utility terms. As such, our analysis represents an extension of Manski (1988), who studies identification of binary choice models without parametric distributional assumptions, to models with social interactions and an extension of Brock and Durlauf (2001a,b), who study identification of social interactions in discrete choice models, to a broader set of environments. Beyond an amalgamation of the analyses in these papers, we also analyze the identification of social interaction in the presence of unobserved group effects.

Section 2 of the paper describes the basic theoretical structure of binary choice social interactions models and discusses econometrically implementable versions of the models. Section 3 provides results on identification when group memberships are randomly assigned. Section 4 discusses identification when group memberships are nonrandom. Section 5 provides conclusions.

2. Binary choice with social interactions

A general model of binary choice with social interactions is developed in Brock and Durlauf (2001a,b) and is the template for our identification analysis. We consider a sample of I individuals; individual i is a member of group g; the group memberships are known to the econometrician. There are $g = 1...G$ groups in the population.

Individual choices are coded by $\omega_i \in \{-1,1\}$. These choices are determined by five factors:

1. observable individual-specific characteristics, measured by an r-vector X_i,

2. unobservable individual characteristics summarized by a scalar ε_i.

3. observable group characteristics, measured by an s-vector Y_g; these are known as contextual effects as they relate to how characteristics of a group affect its members,

4. unobservable (to the econometrician) group characteristics, measured by a scalar α_g,

5. subjective expectation by agent i of the average choice in the group $m^e_{i,g}$; this is known as an endogenous effect as it describes how the behaviors of others (mediated by beliefs) affect each individual.

These factors are assumed to produce payoffs for the possible choices, $u_i(1)$ and $u_i(-1)$ such that the difference between these payoffs is additive in the various factors, i.e.

$$u_i(1) - u_i(-1) = k + cX_i + dY_g + Jm^e_{i,g} + \alpha_g - \varepsilon_i \qquad (1)[3]$$

[3] In eq. (1) and elsewhere, it is understood that variables indexed by g refer to the group of which i is a member whenever the equation describes an individual-level relationship.

Choice 1 is made when this difference is positive. The random payoff terms ε_i are assumed to be drawn from a common distribution function $F_{\varepsilon|X_i,Y_g,\alpha_g}$ and are conditionally (given X_i, Y_g and α_g) independent.

We assume that subjective beliefs are rational, given information on the group level characteristics Y_g and $F_{X|Y_g,\alpha_g}$, the distribution function of X_i conditional on the observable and unobservable (to the econometrician) group characteristics; this last term allows for heteroskedasticity in the distribution of individual characteristics across groups, for example. Hence the subjective expectations $m_{i,g}^e$ coincide with m_g, the mathematical expectation of the average choice in group g given Y_g and α_g. Since

$$E\left(\omega_i \middle| X_i, Y_g, \alpha_g\right) = 2F_{\varepsilon|X_i,Y_g,\alpha_g}\left(k + cX_i + dY_g + Jm_g^e + \alpha_g\right) - 1 \qquad (2)$$

m_g is defined by the integral

$$m_g = 2\int F_{\varepsilon|X,Y_g,\alpha_g}\left(k + cX + dY_g + Jm_g + \alpha_g\right) dF_{X|Y_g,\alpha_g} - 1. \qquad (3)$$

It is possible for there to exist multiple values of m_g that fulfill (3); intuitively the dependence of the payoff (1) on m_g produces an expectations-based complementarity[4] between individual decisions that can generate multiple self-consistent behaviors in the population as a whole. As shown in Brock and Durlauf (2003), multiplicity versus uniqueness in the equilibrium expected average choice levels will depend on the interplay of the strength of social interactions, measured by J and features of the distribution of $F_{\varepsilon|Y_g,\alpha_g}$, specifically the degree of dispersion of the random payoff terms; Theorem 3 of that paper specifically shows that for a given set of parameters k, c, d, J and distribution functions $F_{\varepsilon|Y_g,\alpha_g}$ and $F_{X|Y_g,\alpha_g}$, there will always exist a threshold $\bar{J}$ (which

[4]See Cooper (1999) for a general discussion of the relationship between complementarities and multiple equilibria.

depends on the parameters and distributions), such that if $J > \bar{J}$, there must exist multiple values of m_g that solve (3). Note that when multiple solutions exist, it is assumed that agents know which solution describes expected average choices in g; put differently, agents know which equilibrium expected average choice level is selected.

Our objective in the following discussion is to determine conditions under which information about k, c, d, and J may be obtained from data. We will focus on both point identification, i.e. conditions under which the values of these parameters are identified, as well as partial identification,[5] i.e. conditions under which nontrivial restrictions on the possible values of k, c, d, and J may be obtained. Our discussion will be particularly concerned with the social interactions parameters d and J. The J parameter is of particular interest as it is the hypothesis that $J > 0$ that typically underlies theoretical models of social interactions. In fact, while it is relatively uncontroversial to assume that students in a class are affected by teacher quality or educational resources available in a classroom (both of which are contextual effects) it is more controversial to assume that the effort or learning of one student depends on the effort or learning of others. And as indicated above, from a theoretical perspective, the finding that $J > 0$ is of particular importance because of its implications for the presence of multiple equilibria. From the perspective of econometrics, an important class of identification issues is driven by the question of when contextual social interactions effects generated by Y_g may be distinguished from endogenous social interaction effects generated by m_g. The difficulties associated with distinguishing these two types of social interactions occur because, as indicated by (3), m_g is functionally dependent on Y_g; in a seminal paper Manski (1993) named this the reflection problem.

Identification arguments are naturally based on technical conditions concerning the distributions of errors and regressors. As such, it is often difficult to interpret requirements for identification in terms of substantive restrictions on individuals, and in our case, groups. For this reason, we believe it is useful to have an example in mind. We focus on the analysis of classroom peer group effects: a question addressed by Angrist and Lang (2002), Boozer and Cacciola (2002), Hanushek, Kain, Markman and

Rivkin (2003) and Hoxby (2000a,b) among others. Studies of this type consider the effects of classroom composition on individual educational outcomes. As such, they naturally involve contextual effects such as teacher or school quality as well as endogenous effects produced by peer influences. The classroom example will be useful in assessing the plausibility of various assumptions we make.

There are a number of previous studies of identification. Manski (1993) studies identification in linear and nonparametric contexts and emphasizes the absence of identification when little prior information is available. Brock and Durlauf (2001b,2003) analyze identification for binary and multinomial choice models. They show that for relatively weak conditions on the joint distribution of X_i and Y_g, the parameters of (1) are identified when 1) $F_{\varepsilon|X_i,Y_g,\alpha_g} = F_\varepsilon$, so that the random payoff terms are identicially distributed across all individuals regardless of group membership or individual charactetristics and 2) F_ε is known a priori. The subsequent discussion shows how to extend this to the case where F_ε is unknown and also explores how various types of unobservable variables affect identification. Other contributions to the study of the identification of social interactions include Bayer and Timmins (2002) who analyze identification in multinomial logit models with various types of unobservables.

In evaluating the conditions under which identification does or does not hold for this model, we will follow the standard practice of assuming that population distributions are known for all observables. Hence we ignore issues of estimation.[6]

3. Identification with random assignment and no group level unobservables

In this section, we consider identification for the binary choice model of social interactions when agents are randomly assigned to groups and when there are no group level unobservables present in the individual payoffs defined by (1). Each of these assumptions represents a substantive restriction on the class of environments under study.

---

[5]Manski (2003) provides a definitive analysis of partial identification.
[6]Bisin, Moro, and Topa (2002) discuss estimation strategies for this model.

While random assignment may be appropriate for examples such as classrooms in which schools explicitly follow such policies in forming classes, it clearly does not apply in contexts such as residential neighborhoods; in fact analyses such as Bénabou (1993,1996), Durlauf (1996a,b) and Hoff and Sen (2002) make clear that social interactions are an important factor in understanding neighborhood composition. Similarly, the absence of group level unobservables represents a strong assumption in terms of model specification. In a context such as neighborhoods, a range of social factors such as willingness to contribute to public goods are typically unobservable.[7]

We work with the following set of assumptions in order to provide a baseline for identification analysis. Our strategy is to first understand why identification holds under this particular set of assumptions and then consider how relaxation of random assignment and the absence of group level unobservables affects what may be learned about social interactions. We work with the following assumptions; in the assumptions and elsewhere $\mathrm{supp}\left(Z\right)$ denotes the support of the random vector Z.

$A.1.\ F_{\varepsilon|X_i,Y_g} = F_{\varepsilon};\ F_{\varepsilon}\left(0\right) = .5.$

$A.2.$ F_{ε} is absolutely continuous with associated density $dF_{\varepsilon}; dF_{\varepsilon}$ is positive almost everywhere on the support $\left(L,U\right)$ which may be $\left(-\infty,\infty\right)$.

$A.3.$ X does not include a constant; there exists a group g_0 such that for at least one x_j (with associated nonzero coefficient c_j), x_{ij} varies continuously over R and $\mathrm{supp}\left(X_{-j}\right)$ is not contained in a hyperplane of R^{r-1}.

---

[7]Recent efforts to measure these types of group characteristics are currently under way. A very important example is the Project on Human Development in Chicago Neighborhoods, in which neighborhood characteristics such as collective efficacy, which refers to factors such as how a neighborhood provides support for its members, for example through assistance to neighbors in childrearing are being measured via extensive

$A.4.$ Y does not include a constant; there exists at least one y_l (with associated nonzero coefficient d_l) that varies continuously over R and $\text{supp}\left(Y_{-l}\right)$ is not contained in a hyperplane of R^{s-1}.

$A.5.$ $F_{X|Y_g} = F_X$.

$A.6.$ $\alpha_g = 0 \ \forall g$.

Variants of these assumptions will be maintained throughout and are based on Cameron and Heckman (1998). They are somewhat more restrictive than those that appear in Manski (1988) but are easier to interpret, and so we employ them. Assumption $A.1$ imposes an i.i.d. assumption on the random payoff terms ε_i within and across groups. $A.2$ imposes a certain degree of smoothness on F_ε. $A.3$ and $A.4$ impose linear independence among the observable individual-specific and group-specific characteristics as well as a "large support" assumption on one element of each. $A.5$ operationalizes the idea of random assignment by equating it with the independence of the distribution of individual characteristics within a group from the observable characteristics of the group, i.e. Y_g. $A.6$ eliminates the unobserved group effects.

The set of parameters k, c, d, J and the distribution function F_ε are observationally equivalent to the alternative set of parameters $\bar{k}$, $\bar{c}$, $\bar{d}$, $\bar{J}$ and alternative distribution function $\bar{F}_\varepsilon$ if

$$F_\varepsilon\left(k + cX_i + dY_g + Jm_g\right) = \bar{F}_\varepsilon\left(\bar{k} + \bar{c}X_i + \bar{d}Y_g + \bar{J}m_g\right) \ \forall i \tag{4}$$

and

survey efforts. Important studies based on this project include Sampson, Raudenbush and Earls (1997) and Sampson, Morenoff, and Earls (1999).

$$m_g = 2\int F_\varepsilon\left(k + cX + dY_g + Jm_g\right)dF_X - 1 = 2\int \overline{F}_\varepsilon\left(\overline{k} + \overline{c}X + \overline{d}Y_g + \overline{J}m_g\right)dF_X - 1 \quad (5)$$

for all elements of $\mathrm{supp}(X)$ and $\mathrm{supp}(Y)$. Identification holds if observational equivalence between the model parameters and distribution function and an alternative implies they are identical, i.e.

$$k = \overline{k},\ \mathrm{c} = \overline{c},\ d = \overline{d},\ J = \overline{J} \text{ and } F_\varepsilon = \overline{F}_\varepsilon. \quad (6)$$

Notice that in the conditions for observational equivalence, eqs. (4) and (5) we assume that m_g is observable since it is a moment associated with observables.

Assumptions A.1-A.6 provide our first identification result:

Proposition 1: Identification of the binary choice model with social interactions and random assignment to groups

The parameters of the binary choice model with social interactions (1) under self-consistency condition (3) are identified up to scale under assumptions $A.1 - A.6$.

Proof. Relative to Manski (1988), the key to establishing identification in the presence of social interactions concerns the treatment of m_g. Put differently, if it were known that $J = 0$, so that m_g did not affect individual choices, then identification of the remaining parameters up to scale has been shown in Manski (1988) Proposition 2, Corollary 5. The key features of Manski's proof are 1) the quantile independence of $F_{\varepsilon|Y}$, i.e. $F_{\varepsilon|Y_g}(z) = F_\varepsilon(z)$ for some z and all g, which is implied by A.1 and 2) the large support assumption on x_1, which corresponds to A.3. The regressor m_g, in turn, is a function of F_ε, F_X, Y_g and (if there is more than one solution to (3)) the selection rule for m_g.[8] The extension of Manski's results to the social interactions case requires that the endogeneity of m_g under (3) is accounted for.

First, we consider identification of c and F_ε. Without loss of generality, assume that x_1 is the element of X with nonzero coefficient and continuous variation across R; we normalize $c_1 = 1$. For individuals in group g_0, let $\kappa_{g_0} = k + dY_{g_0} + Jm_{g_0}$ and $\bar{\kappa}_{g_0} = \bar{k} + \bar{d}Y_{g_0} + \bar{J}m_{g_0}$. Hence for members of the group, the probability that $\omega_i = 1$ is $F_\varepsilon\left(\kappa_{g_0} + cX_i\right)$ under the true set of parameters and random payoff distribution function; under the alternative this probability is $\bar{F}_\varepsilon\left(\bar{\kappa}_{g_0} + \bar{c}X_i\right)$.

Identification of c and F requires that for all elements of $\operatorname{supp}\left(X\right)$ among members of g_0,

$$\begin{gathered} F_\varepsilon\left(\kappa_{g_0} + cX_i\right) = \bar{F}_\varepsilon\left(\bar{\kappa}_{g_0} + \bar{c}X_i\right) \ \forall X_i \in \operatorname{supp}\left(X\right) \Rightarrow \\ c = \bar{c},\ F_\varepsilon = \bar{F}_\varepsilon \end{gathered} \tag{7}$$

Identification in the sense of (4) is established in Manski (1988) Proposition 2, Corollary 5. Therefore, the analysis of identification of the remaining model parameters may be conducted under the assumption that c and F_ε are known.

We consider the identification of the remaining parameters. Using the random assignment assumption *A*.5, define a mapping $\psi(\cdot)$ that is invariant across groups[9]

[8]From the perspective of identification, it is not necessary to model how an equilibrium is selected for a given group, since m_g is observable as it is a sample moment (as discussed in the text). However, the issue of selection does matter for estimation, since the estimation implicitly requires construction of an estimate of m_g that is consistent with the estimated parameters. Some suggestions in this respect are found in Bisin, Moro, and Topa (2002); this is a problem we are currently working on as well.

[9]Invariance of $\psi(\cdot)$ follows from the assumptions $A.1$ and $A.5$ which state that neither F_ε nor F_X varies across groups.

$$\psi(\varsigma) = 2\int F_{\varepsilon}(cX + \varsigma)\, dF_X - 1 \tag{8}$$

It is obvious that $\psi(\cdot)$ is monotone increasing and nonlinear in ς. Taking $\varsigma = k + dY_g + Jm_g$, then

$$\psi(k + dY_g + Jm_g) = 2\int F_{\varepsilon}(k + cX + dY_g + Jm_g)\, dF_X - 1 = m_g \tag{9}$$

where the second equality in (9) restates (3).

This mapping facilitates identification analysis. Monotonicity of $\psi(\cdot)$ implies that the second observational equivalence requirement (eq. (5)) holds if and only if

$$k + dY_g + Jm_g = \bar{k} + \bar{d}Y_g + \bar{J}m_g \ \ \forall Y_g \in \operatorname{supp}(Y) \tag{10}$$

Rewriting (10) as

$$\left(J - \bar{J}\right) m_g = \bar{k} - k + \left(\bar{d} - d\right) Y_g \ \ \forall Y_g \in \operatorname{supp}(Y) \tag{11}$$

it is apparent that observational equivalence between two distinct sets of parameters requires that m_g is a linear function of Y_g, since by $A.4$ at least one element of Y_g varies continuously across R. However, such a linear relationship cannot exist since m_g is bounded between -1 and 1. Therefore, (11) can only hold if $J = \bar{J}$. Therefore J is identified.

Imposing $J = \bar{J}$, observational equivalence of k and d with $\bar{k}$ and $\bar{d}$ requires that

$$\bar{k} - k + \left(\bar{d} - d\right) Y_g = 0 \ \forall Y_g \in \operatorname{supp}(Y) \tag{12}$$

Under the assumption $A.4$, this can only hold if $k=\overline{k}$ and $d=\overline{d}$ which completes the proof that the model parameters are k, c, d and J.

What are the key features of the proof? Assumptions $A.1$- $A.3$ allow one to identify c and F using intra-group data from one group. As such, the assumptions do nothing more than allow one to invoke Manski (1988) Proposition 2, Corollary 5. Relative to the social interactions model, the conditions allow one to construct m_g across neighborhoods and therefore treat it as an observable, given a set of parameters k, c, d and J. Assumptions $A.5$ and $A.6$ mean that the self-consistency condition (5) for observational equivalence of expected average choice levels may be inverted to produce a condition for observational equivalence on the linear payoff differentials between the choices, i.e. (4). Assumption $A.4$ ensures that there is enough variability in Y_g to identify k, d, and J given this transformation of the integral condition (5) into the linear condition (10).

The identification argument exploits the fact that because m_g is bounded between -1 and 1, it cannot be linearly dependent on Y_g under the large support assumption $A.4$. To be clear, what drives our result is not that Y_g has unbounded support, but rather that Y_g possesses a sufficiently large support to reveal the nonlinear relationship between Y_g and m_g that holds for any given set of parameters and a given random payoff distribution, cf. eq. (3) and its form under random assignment, eq. (5). This is the reason why the so-called reflection problem (Manski (1993)) does not arise in the binary choice case; see Brock and Durlauf (2001a,b) for more discussion. The large support assumption is a simple way of ensuring that the data reveal the nonlinear relationship between m_g and Y_g. To understand the importance of this nonlinearity, it is useful to recall Manski's (1993) nonidentification result for linear models. Manski considers the model

$$\omega_i = k + cX_i + dY_g + Jm_g^e + \varepsilon_i \tag{13}$$

under the assumption that $Y_g = E\left(X_i \middle| Y_g\right)$, i.e. the contextual variables are the group level averages of the individual characteristics. Rationality of subjective beliefs, using the same information structure as used to derive (3) implies that

$$m_g = \frac{k}{1-J} + \frac{c+d}{1-J} Y_g \tag{14}$$

so that m_g is linearly dependent on the other regressors in (13). This linear dependence is ruled out in binary choice models.[10] [11]

This analysis indicates that the unobservability of dF_ε is not an impediment to identification; however, it is important to note that accurate estimates of the magnitude of social interactions will critically depend on the accuracy of estimates of this density. Our proof relies on the identifiability of dF_ε using observations within a single group. In a context such as classrooms, it may be problematic to estimate dF_ε using data from a single group. In such cases, it will be important to employ estimation methods that allow information on dF_ε to be constructed from data from all groups.

4. Group level unobservables

As we have suggested, one important limitation to Proposition 1 is that it assumes that no group level unobservables are present in the individual payoff functions. Such unobservables are likely in many contexts even when there is random assignment. In the case of classrooms, differences in teacher quality are one such unobservable. In this section, we relax assumption $A.6$ and consider some ways to achieve identification when

[10]The boundedness of choices is important in establishing identification in other contexts. For example, it plays a key role in the development of bounds on treatment effects, cf. Manski (1990).

[11]This same nonlinearity argument also implies that parametric multinomial choice models are identified when group memberships are randomly assigned and when there are no group level unobservables; see Brock and Durlauf (2003) for a formal statement and proof.

group-level unobservables are present. The analysis will show that prior information on the distribution of group level unobservables is necessary to achieve even partial identification of social interactions; we provide examples of prior information that is useful in this respect.[12]

In order to preserve the random assignment logic, we further replace $A.1$ and $A.5$ with $A.1'$ and $A.5'$

$$A.1'.\ F_{\varepsilon|X_i,Y_g,\alpha_g} = F_{\varepsilon};\ F_{\varepsilon}(0) = .5.$$

$$A.5'.\ F_{X|Y_g,\alpha_g} = F_X\ .$$

These modified assumptions preserve the equality across groups of the intragroup distributions of individual observable and unobservable terms.

When group level unobservables are present, our identification definition needs to account for their presence. Modifying eqs. (4) and (5), identification in the presence of unobservable group effects requires that for a given set of parameter values $k,\ c,\ d,\ J$, set of unobservables α_g, and random payoff distribution F_{ε}, that if there exists an alternative set of parameters $\bar{k},\ \bar{c},\ \bar{d},\ \bar{J}$, set of unobservables $\bar{\alpha}_g$, and distribution function $\bar{F}_{\varepsilon}$ such that

$$F_{\varepsilon}\left(k + cX_i + dY_g + Jm_g + \alpha_g\right) = \bar{F}_{\varepsilon}\left(\bar{k} + \bar{c}X_i + \bar{d}Y_g + \bar{J}m_g + \bar{\alpha}_g\right)\ \forall i \qquad (15)$$

and

[12]When individual agents face more than two choices, additional routes to identification may exist beyond those we study, as shown in Bayer and Timmins (2002); the basic idea in this work is that comparisons among subsets of the choices may be used to identify unobservable choice-specific payoff terms.

$$\begin{aligned} m_g &= 2\int F_\varepsilon\left(k + cX + dY_g + Jm_g + \alpha_g\right)dF_X - 1 \\ &= 2\int \bar{F}_\varepsilon\left(\bar{k} + \bar{c}X + \bar{d}Y_g + \bar{J}m_g + \bar{\alpha}_g\right)dF_X - 1 \end{aligned} \tag{16}$$

for all elements of $\text{supp}(X)$ and $\text{supp}(Y)$, it must be the case that eq. (6) holds. Relative to the earlier definition of identification, notice that the values of the unobservables, i.e. α_g and $\bar{\alpha}_g$ must be allowed to differ when comparing alternative sets of parameters.

Without any additional assumptions, identification fails for the social interactions model with unobserved group effects. This is easy to see. Clearly, identification of c and dF_ε is unaffected by the presence of α_g; the argument in Section 3 is unaffected since in the current case all that changes is that $\kappa_{g_0} = k + dY_{g_0} + \alpha_{g_0} + Jm_{g_0}$. Hence, we proceed taking these as known.

We again work with the mapping $\psi(\varsigma)$ defined by equation (8). For $\varsigma = k + dY_g + Jm_g + \alpha_g$,

$$\psi\left(k + dY_g + Jm_g + \alpha_g\right) = 2\int F_\varepsilon\left(k + cX + dY_g + Jm_g + \alpha_g\right)dF_X - 1 = m_g \tag{17}$$

Comparing (17) and (9), observational equivalence requires, recalling that $\psi(\varsigma)$ is increasing, monotonic and invariant across groups,

$$k + dY_g + Jm_g + \alpha_g = \bar{k} + \bar{d}Y_g + \bar{J}m_g + \bar{\alpha}_g \quad \forall Y_g \in \text{supp}(Y) \tag{18}$$

If one chooses $\bar{\alpha}_g = \alpha_g + Jm_g$ and $\bar{J} = 0$, then $k + dY_g + Jm_g + \alpha_g = k + \bar{d}Y_g + \bar{\alpha}_g$ $\forall Y_g \in \text{supp}(Y)$, which means that J and d are not identified. We can therefore state

Proposition 2. Nonidentification with unobserved contextual effects

The parameters of the binary choice model with social interactions (1) with self-consistency condition 3, are not identified under assumptions $A.1'$, $A.2$, $A.3$, $A.4$ and $A.5'$.

What sorts of additional assumptions beyond $A.1'$, $A.2$, $A.3$, $A.4$ and $A.5'$ will imply identification? A number of possibilities exist that may, depending on the particular environment, be plausible. We consider two alternative classes of identification strategies.

i. identification based on restrictions of the joint distribution of observables and unobservables

One way in which identification may be achieved in the presence of unobservables is via the use of prior restrictions on the distribution of these unobservables either in isolation or in terms of their relationship to various observables. The utility of this approach will of course depend on the plausibility of these restrictions.

a. restrictions on the support of the unobservables

In the proof of Proposition 1 a large support assumption, combined with the intrinsic nonlinearity of the binary choice model, produced identification. We first explore how the logic of the large support assumption may be adapted to the model with group-level unobservables. An analogous large support assumption in the unobservable case is

U.1. $\text{supp}\left(\left(d-\bar{d}\right)Y_g+\alpha_g-\bar{\alpha}_g\right)$ is unbounded.

This assumption produces identification of J but does not permit identification of d, as stated in Proposition 3.

Proposition 3. Partial identification of the binary choice model with social interactions and unobservables with unbounded support

The parameter J is identified up to scale but k and d are not identified under assumptions $A.1'$, $A.2$, $A.3$, $A.4$, $A.5'$ and $U.1$.

Proof. Rewrite (18) as

$$\left(\bar{J}-J\right)m_g = k-\bar{k}+\left(d-\bar{d}\right)Y_g+\alpha_g-\bar{\alpha}_g \tag{19}$$

Under $U.1$, the right hand side of (19) will possess full support. Since $m_g \in \left[-1,1\right]$, (19) cannot hold unless $J=\bar{J}$, so the J is identified. Imposing this on (19),

$$k-\bar{k}+\left(d-\bar{d}\right)Y_g+\alpha_g-\bar{\alpha}_g=0 \tag{20}$$

which imposes no restrictions on k and d since α_g and $\bar{\alpha}_g$ are unrestricted. This verifies the proposition. $\square$

In our judgment, this is a relatively unappealing route to identification since it is difficult to fairly imagine a priori cases in which the unbounded support assumption $U.1$ can be made credible. Perhaps one could identify cases where, for example, government policies of some type have introduced large unmeasured heterogeneity into schools via lumpy investments in facilities.

b. restrictions on the relationship between observables and unobservables

A second approach to restricting the unobservable contextual effects so as to achieve some sort of identification is through restrictions on the relationship between the unobservables and observable contextual effects. We consider two sorts of restrictions of this type.

stochastic monotonicity

One way to restrict the relationship between observables and unobservables is to impose stochastic monotonicity between the types of contextual effects. This type of assumption is analogous to those studied in Manski (1997) and Manski and Pepper (2000). Without loss of generality, assume that each element of Y_g is measured so that $d \geq 0$[13] with strict inequality for at least one element of d. We consider the following assumption:

M.1. First order stochastic monotonicity of group level unobservables. If $Y_g > Y_{g'}$, then the conditional distribution of unobservables in g', $F_{\alpha_{g'}|Y_{g'}}$, is first order stochastically dominated by $F_{\alpha_g|Y_g}$.

This assumption is sufficient to facilitate partial identification of social interactions.

Proposition 4. Pattern reversals and partial identification of endogenous social interactions

If there exists a pair of groups such that

$$Y_g > Y_{g'} \text{ and } E\left(m_g \middle| Y_g\right) < E\left(m_{g'} \middle| Y_{g'}\right) \tag{21}$$

then it must be the case that $J > 0$ and J is large enough to produce multiple equilibria under assumptions $A.1'$, $A.2$, $A.3$, $A.4$, $A.5'$ and $M.1$.

Proof. From (3), $E\left(m_g \middle| Y_g\right)$ is defined by

$$E\left(m_g \middle| Y_g\right) = 2\int F_\varepsilon\left(k + cX + dY_g + Jm_g + \alpha_g\right) dF_X\, dF_{\alpha_g|Y_g} - 1 \qquad (22)$$

If the observables for group g are ordered according to $Y_g > Y_{g'}$ so that by $M.1$ $F_{\alpha_g|Y_g}$ first-order stochastically dominates $F_{\alpha_{g'}|Y_{g'}}$, and if $J = 0$, then $E\left(m_g \middle| Y_g\right) > E\left(m_{g'} \middle| Y_{g'}\right)$ since $F_\varepsilon\left(k + cX + dY_g + Jm_g + \alpha_g\right)$ is increasing in Y_g (recall that Y_g is measured so that $d \geq 0$) and α_g. This monotonicity can only be broken if $J > 0$ and there are multiple equilibria, so that the equilibrium expected average choice levels of groups g and g' are such that g has coordinated at an equilibrium other than the one with the highest expected average choice level where g' has coordinated at an equilibrium other than the lowest expected average choice; otherwise the pattern reversal in (21) could not have occurred. □

One may consider alternatives to $M.1$; one natural example is stochastic dominance based on an index, i.e.

$M.2$. First order stochastic monotonicity of an index of unobservables. There exists a scalar index function $I(\ )$ such that if $I\left(Y_g\right) > I\left(Y_{g'}\right)$, then the conditional distribution of unobservables in g', $F_{\alpha_{g'}|I\left(Y_{g'}\right)}$, is first order stochastically dominated by $F_{\alpha_g|I\left(Y_g\right)}$. This allows one to state a Corollary to Proposition 4.

Corollary 1. Pattern reversals with an index and partial identification of social interactions

If there exists a pair of groups such that

---

[13]Vector inequalities are measured in the usual way, i.e. $V_1 > V_2$ means each element of V_1 is greater than the corresponding element of V_2.

$$I\left(Y_g\right) > I\left(Y_{g'}\right) \text{ and } E\left(m_g \left| I\left(Y_g\right)\right.\right) < E\left(m_{g'} \left| I\left(Y_{g'}\right)\right.\right) \tag{23}$$

then it must be the case that $J > 0$ and J is large enough to produce multiple equilibria under assumptions $A.1'$, $A.2$, $A.3$, $A.4$, $A.5'$ and $M.2$.

This corollary includes cases such as $I\left(Y_g\right) = dY_g$.

Are monotonicity assumptions plausible? This is of course a matter of judgment; nevertheless we believe there are cases where this is so. For example, consider the case of teenage smoking. According to one recent national survey in 2003, 10.0% of black teenagers had smoked a cigarette in the last 30 days whereas 29.4% of white teenagers had. As ethnicity is a natural social group, this sort of behavioral difference seems hard to understand without endogenous social interactions, since one would generally expect that the contextual factors that are omitted by these raw percentages, e.g family income, etc. would have the property of making the black teenage smoking rate higher than the white rate.[14] Similarly, we could conjecture that for unobservable teacher quality, stochastic monotonicity might hold with respect to observable variables such as average family income or per pupil expenditure.[15]

linearity

Alternatively, one can assume a linear relationship between the observable and unobservable contextual effects.

L.1. $\alpha_g = f_0 + f_1 Y_g + \mu_g$, $E\left(\mu_g \left| Y_g\right.\right) = 0$

---

[14]Various social interaction effects have been directly estimated in studies such as Krauth (2002).

[15]The power of the stochastic monotonicity assumption also implies the importance of testing it when possible. Barrett and Donald (2003) have recently proposed methods for doing this that may be relevant in this context.

This is a much stronger assumption than stochastic monotonicity, but will in turn provide an important increase in the restrictions that $J = 0$ places on data. To see this, we construct the variable $Z_i = F_\varepsilon^{-1}\left(\text{Prob}\left(\omega_i = 1 \middle| X_i, Y_g, \alpha_g\right)\right)$ which by (1) equals

$$Z_i = k + cX_i + dY_g + Jm_g + \alpha_g \tag{24}$$

One can compare this to the regression equation

$$Z_i = \pi_0 + \pi_1 X_i + \pi_2 Y_g + \phi_i \tag{25}$$

Under $J = 0$ and assumption $L.1$, $E\left(\phi_i \middle| X_i, Y_g\right) = 0$ whereas if $J \neq 0$ there is a nonlinear dependence between ϕ_i and X_i and Y_g, given the nonlinearity inherent in the relationship between these vectors and m_g as shown in (3). Assumption *L.1* thus places a restriction on Z_i.

Proposition 5. Nonlinearity of transformed outcome probabilities in the presence of endogenous social interactions

If $J = 0$**,** then Z_i is linear in X_i and Y_g whereas if $J \neq 0$, then Z_i is not linear in X_i and Y_g, under assumptions $A.1'$, $A.2$, $A.3$, $A.4$, $A.5'$ and $L.1$.

This implication is easily testable if one can construct an empirical analog to Z_i and so estimate (25) as a regression. Granger and Terasvirta (1996) provide methods for testing whether the linear relationship in (25) holds. One can also use the BDS test (Brock, Dechert, LeBaron, and Scheinkman (1996)) on estimated residuals from (25) to see whether the residuals are i.i.d.; acceptance of this null rules out the nonlinearity described in the Proposition.

Can one construct such an empirical analog? One possibility is suggested by the classroom example. Suppose that students answer a number of true false questions and that the proportion for student i is $\hat{s}_i$. Assume that the probability of a correct answer follows (1), i.e. the probability of a given answer being correct is $F_\varepsilon\left(k+cX_i+dY_g+Jm_g\right)$. If $\hat{F}_\varepsilon$ is an estimate of F_ε, which is identified under our assumptions, then $\hat{F}_\varepsilon^{-1}\left(\hat{s}_i\right)$ is a sample estimate of Z_i.

restrictions on the density of unobservables

A third route to identification in the presence of unobserved group effects lies in restricting the conditional density of the unobserved group effects given observed group characteristics, i.e. $dF_{\alpha_g|Y_g}$. One such restriction is

$P.1$. $dF_{\alpha_g|Y_g}$ is unimodal for all Y_g.

This assumption leads to Proposition 6.

Proposition 6. Partial identification of endogenous social interactions when unobservables are unimodally distributed.

If $J=0$, then $dF_{m_g|Y_g}$ is unimodal whereas if $dF_{m_g|Y_g}$ is multimodal, then $J>0$ and J is large enough to produce multiple equilibria, under assumptions $A.1'$, $A.2$, $A.3$, $A.4$, $A.5'$, and $P.1$.

Proof. Recall from (17) that $m_g=\psi\left(k+dY_g+\alpha_g\right)$. For fixed Y_g, one can compute the conditional density $dF_{m_g|Y_g}$ using ψ since ψ is monotonic and increasing via

$$\begin{gathered}dF_{m_g|Y_g}\left(\xi\right)=\Pr\left(m_g=\xi\,\middle|\,Y_g\right)=\\ \Pr\left(\psi\left(k+dY_g+\alpha_g\right)=\xi\,\middle|\,Y_g\right)=\Pr\left(k+dY_g+\alpha_g=\psi^{-1}\left(\xi\right)\,\middle|\,Y_g\right)=\\ \Pr\left(\alpha_g=\psi^{-1}\left(\xi\right)-k-dY_g\,\middle|\,Y_g\right)=dF_{\alpha_g|Y_g}\left(\psi^{-1}\left(\xi\right)-k-dY_g\right)\end{gathered}\tag{26}$$

Given the unimodality assumption $P.1$, let α^* denote the maximum of $dF_{\alpha_g|Y_g}$ and $m^* = \psi\left(k + dY_g + \alpha^*\right)$. Suppose that m^* is not the unique maximum of $dF_{m_g|Y_g}$. If not, then there exists a m^{**} such that $dF_{m_g|Y_g}\left(m^{**}\right) > dF_{m_g|Y_g}\left(m^*\right)$. But this means that for α^{**} defined by $m^{**} = \psi\left(k + dY_g + \alpha^{**}\right)$ it must be the case that $dF_{\alpha_g|Y_g}\left(\alpha^{**}\right) > dF_{\alpha_g|Y_g}\left(\alpha^*\right)$, which contradicts the definition of α^* and verifies the proposition. □

The plausibility of unimodality will of course depend on context. We conjecture that for a variable such as unobserved teacher quality, the assumption of unimodality is reasonable. In applying this condition, it is important to recognize that multimodality of $dF_{m_g|Y_g}$ does not imply multimodality of dF_{m_g}. Hence the arguments against multiple equilibria used by Glaeser, Sacerdote, and Scheinkman (1996) which are in essence based on the unimodality of dF_{m_g} do not apply to this context.[16]

ii. identification in panels

Our discussion thus far has focused on identification in cross-section contexts. An alternative possibility for addressing group level unobservables lies in the use of panel data. In the linear context, panels have been shown to affect identification in Brock and Durlauf (2001b) and Graham and Hahn (2003). In the linear context, panels facilitate identification for two reasons. First, differencing of the data permits elimination of group-level fixed effects. Second, the timing of contextual and endogenous social interactions (eg. whether $\omega_{i,t}$ depends on $Y_{g,t}$ and $m_{g,t}$ or $Y_{g,t-1}$ and $m_{g,t-1}$) can break the linear dependence between the contextual and endogenous effects that is the basis of Manski's

[16] The reason why here and elsewhere multiple equilibria do not imply multimodality of dF_{m_g} is that dF_{m_g} in essence represents a mixture distribution where the different equilibria constitute components of a mixture. Mixture distributions are not necessarily multimodal; see Lindsay (1995, pg. 4-5) for a nice example.

(1993) finding of nonidentification in the linear case. We will focus on the analog of differencing to see how the availability of panel data can affect identification; as such, our argument is an example of the analysis pioneered in Chamberlain (1984).

Implications of time varying group membership characteristics

Identification may be facilitated if group characteristics evolve across time. This is the key idea used to identify peer effects in Hoxby (2000b); here we provide a formalization and generalization of her argument. To do this, we replace (1) with

$$u_i(1) - u_i(-1) = k + cX_i + dY_g + eY_{g,t} + Jm_{g,t} + \alpha_g - \varepsilon_{i,t} \tag{27}$$

where, relative to (1), Y_g and $Y_{g,t}$ are vectors of group-specific characteristics that do and do not vary across time. In the context of classroom interactions, we are interpreting Y_g as incorporating variables such as observed teacher quality, for example, whereas we are interpreting $Y_{g,t}$ as incorporating variables such as the average parental income. In this context, following previous arguments, one can construct a map $\psi_t(\varsigma)$

$$\psi_t(\varsigma) = 2\int F_\varepsilon(cX + \varsigma)\, dF_{X_t} - 1 \tag{28}$$

In (28), F_{X_t} is the density of X at time t. We are therefore assuming that the distribution of $\varepsilon_{i,t}$ is time invariant but are allowing the distribution $X_{i,t}$ to change temporally. As before, $\psi_t(\varsigma)$ is monotonic increasing and invariant across *g*; the only difference relative to (17) is that the map is allowed to vary across t. In parallel to our earlier discussion, $\varsigma = k + dY_g + eY_{g,t} + Jm_{g,t} + \alpha_g$, producing

$$\begin{gathered}\psi_t\left(k + dY_g + eY_{g,t} + Jm_{g,t} + \alpha_g\right) = \\ 2\int F_\varepsilon\left(k + cX + dY_g + eY_{g,t} + Jm_{g,t} + \alpha_g\right) dF_{X_t} - 1 = m_{g,t}\end{gathered} \tag{29}$$

The invariance of $\psi_t(\varsigma)$ across groups means that observational equivalence requires that for each time t

$$k + dY_g + eY_{g,t} + Jm_{g,t} + \alpha_g = \bar{k} + \bar{d}Y_g + \bar{e}Y_{g,t} + \bar{J}m_{g,t} + \bar{\alpha}_g \qquad (30)$$

which provides a linear panel structure when we consider (30) at different t's.

The panel structure of (30) allows one to achieve identification. Formally, we have Proposition 7.

Proposition 7. Partial identification in binary choice models with panel data

For the model (27) and self-consistency condition (29), c, e, and J are identified whereas k and d are not identified, under assumptions $A.1'$, $A.2$, $A.3$, $A.4$ and $A.5'$.

Proof. c is identified as shown in Proposition 1. To analyze the remaining parameters, difference (30) to produce

$$(e-\bar{e})\left(Y_{g,t} - Y_{g,t-1}\right) + \left(J - \bar{J}\right)\left(m_{g,t} - m_{g,t-1}\right) = 0 \qquad (31)$$

Eq. (31) implies that the $m_{g,t} - m_{g,t-1}$ must be linear for all elements of the support of $Y_{g,t} - Y_{g,t-1}$. If the domain of this support is large enough, then our nonlinearity argument will apply, so that $e = \bar{e}$ and $J = \bar{J}$ and so these parameters are identified. The nonidentifiability of k and d is immediate from the absence of any restrictions on the cross-second moments of the vectors 1 and Y_g with α_g. □

The logic of this proposition may be extended to more alternative frameworks. Hoxby's (2000b) analysis provides a variant of this model in which it is assumed that within group g there are two individual types $T \in \{B, W\}$ such that within-group social interactions that are time varying only occur among members of a common type. Thus, she works with a model of the form

$$u_i(1)-u_i(-1)=k+cX_i+dY_g+eY_{T,g,t}+Jm_{T,g,t}+\alpha_g-\varepsilon_{i,t} \tag{32}$$

The specific idea in her work is that $Y_{T,g,t}$ measures the percentage of T in the classroom; intuitively, uses fluctuations in the percentage of blacks in a classroom to identify social interactions. Her approach suggests a strategy for identifying social interactions that exploits the two dimensional nature of groups at a point in time. Suppose that one observes two types within a classroom. This will allow the construction of the difference between two means of the two groups, $m_{W,g,t}-m_{B,g,t}$ which will prove useful in controlling for the common group effects Y_g and α_g. To see why, note first that for (32), the required self-consistency condition for type-specific expected average choices is

$$m_{T,g,t}=2\int F_{\varepsilon}\left(k+cX_i+dY_g+eY_{T,g,t}+Jm_{T,g,t}+\alpha_g\right)dF_{X|Y_g,Y_{T,g,t}}-1,\ T\in\{B,W\} \tag{33}$$

By analogy to (29) and (30), (33) implies that

$$k+dY_g+eY_{T,g,t}+Jm_{T,g,t}+\alpha_g=\bar{k}+\bar{d}Y_g+\bar{e}Y_{T,g,t}+\bar{J}m_{g,t}+\bar{\alpha}_g \tag{34}$$

so that if one subtracts (34) with $T=W$ from (34) with $T=B$

$$\left(e-\bar{e}\right)\left(Y_{W,g,t}-Y_{B,g,t}\right)+\left(J-\bar{J}\right)\left(m_{W,g,t}-m_{B,g,t}\right)=0 \tag{35}$$

eliminating Y_g and α_g. If there is sufficient variation in $Y_{W,g,t}-Y_{B,g,t}$ across groups, then the nonlinear relationship between $Y_{W,g,t}-Y_{B,g,t}$ and $m_{W,g,t}-m_{B,g,t}$ will require

$e-\overline{e}=J-\overline{J}=0$. This suggests how memberships in multiple groups (e.g. classrooms and race) can facilitate identification.

5. Identification with nonrandom assignment

i. Assignment based on observables

We first consider the case where agents are nonrandomly assigned, but where the assignment is a function of observable variables. What this means is that assumption $A.5$ no longer holds, i.e. different groups will be associated with different distributions of observables $F_{X|Y_g}$. We will preserve the independence of this distribution from the group identity and so replace $A.5$ with $A.5''$

$A.5''$. $F_{X|Y_g,\alpha_g}=F_{X|Y_g}$ and is independent of g.

We will also maintain assumption $A.6$ that there are no unobservable group level variables; this may be replaced with a version of $M.1$; we omit doing this for expositional purposes.

If we maintain assumption $A.1$ that the conditional distribution of the individual-specific random payoff terms is independent of the group characteristics, our analysis would correspond to the case of strong ignorability in the treatment effects literature, cf. Rosenbaum and Rubin (1983).[17] This will not be necessary in what follows; we will work with

[17]In Rosenbaum and Rubin (1983), even though different individuals have different treatment probabilities, the conditional distribution of the effects of treatment is independent of whether treatment occurs, conditional on those variables that determine the treatment probability. In modeling the dependence of group memberships on observables, we are allowing memberships to depend on the observable Y_g, but the median of the random term in the payoff function is independent of these variables.

$A.1''$. $F_{\varepsilon|Y_g,\alpha_g} = F_{\varepsilon|Y_g}$ and is independent of g ; $F_{\varepsilon|Y_g}(0) = .5$.

This assumption means that the group characteristics do not affect the median of the random terms but may affect other aspects of the distribution.

Relative to Proposition 1, selection on observables (as we have defined it), requires additional analysis because the argument that the $\psi(\cdot)$ mapping defined by (8) is invariant across groups is no longer valid. The original argument can be amended, however. Consider the conditional mapping

$$\psi\left(\varsigma \middle| Y_g\right) = 2\int F_{\varepsilon|Y_g}\left(cX + \varsigma\right) dF_{X|Y_g} - 1 \tag{36}$$

For a fixed value of Y_g, $\psi\left(\cdot \middle| Y_g\right)$ is monotonic, increasing, and invariant across groups just as $\psi(\cdot)$ was in the analysis of Proposition 1. This means that if we set $\varsigma = k + dY_g + m_g$, observational equivalence between parameters k, c, d, J and random payoff distribution $F_{\varepsilon|Y_g}$ and alternative parameters $\bar{k}, \bar{c}, \bar{d}, \bar{J}$, and distribution function $\bar{F}_{\varepsilon|Y_g}$ will require that (4) and

$$m_g = 2\int F_{\varepsilon|Y_g}\left(k + cX + dY_g + Jm_g\right) dF_{X|Y_g} - 1 = 2\int \bar{F}_{\varepsilon|Y_g}\left(\bar{k} + \bar{c}X + \bar{d}Y_g + \bar{J}m_g\right) dF_{X|Y_g} - 1 \tag{37}$$

imply (6). The argument made earlier that this is true will again apply in this case. Therefore, we can conclude with this Proposition.

Proposition 8. Identification with selection on observables

The parameters of the binary choice model (1) with self-consistency condition (3) are identified up to scale under assumptions $A.1''$, $A.2$, $A.3$, $A.4$, $A.5''$ and $A.6$.

ii. assignment based on unobservables

We finally consider identification when group memberships are related to unobservable individual characteristics. To do this, we treat the membership question as the outcome of a matching problem and place some restrictions on the equilibria that emerge from the matching.[18] We assume that matching occurs with respect to indices A_i and T_g respectively. We assume that

$$A_i = cX_i - \varepsilon_i \tag{38}$$

and

$$T_g = dY_g \tag{39}$$

In the context of peer effects in classrooms, A_i may be thought of as student ability and T_g as teacher quality.

We assume that matching occurs on the basis of individual and group quality. To do this, we assume that the individual characteristics X_i are measured so that $c \geq 0$. Individuals and groups are matched in the sense that higher group quality is associated with higher individual quality in the sense of the following assumption:

S.1. For any two groups g and g', if $T_g > T_{g'}$, then $F_{A|T_g}$ first order stochastically dominates $F_{A|T_{g'}}$.

[18]This strategy is particularly useful when instrumental variables approaches to accounting for self-selection are not available. See Evans, Oates, and Schwab (1992) for an example. See also Ioannides and Zabel (2003b) for an interesting analysis of how

This assumption is weaker than one which imposes strict assortative matching between better groups and higher ability individuals; the latter is predicted by models such as Becker (1973). The assumption is qualitatively consistent with a range of payoff functions that relate groups and individuals, see Sattinger (1993) for a useful survey of equilibrium matching problems. Notice that the assumption replaces $A.1''$ since it places an implicit restriction on $F_{\varepsilon|Y_g}$

This assumption leads to the following proposition.

Proposition 9. Partial identification of endogenous social interactions under assortative matching

If $T_g > T_{g'}$ and $J = 0$, then $E\left(m_g \middle| T_g\right) > E\left(m_{g'} \middle| T_{g'}\right)$ under assumptions $A.2,\ A.3,\ A.4,\ A.5'',\ A.6$ and *S.1*.

Proof. Consider the binary choice model (1) under the assumption $J = 0$. It is sufficient to show that $\text{Prob}\left(A_i + T_g > 0 \middle| T_g\right)$ is increasing in T_g, since $E\left(m_g \middle| T_g\right) = E\left(\omega_i \middle| T_g\right)$ and $E\left(\omega_i \middle| T_g\right) = 2\text{Prob}\left(A_i + T_g > 0 \middle| T_g\right) - 1$. Rewriting this probability as

$$\text{Prob}\left(A_i + T_g > 0 \middle| T_g\right) = \text{Prob}\left(A_i > -T_g \middle| T_g\right) = 1 - F_{A|T_g}\left(-T_g\right) \tag{40}$$

It is immediate from $S.1$ that $F_{A|T_g}\left(-T_g\right)$ is decreasing in T_g hence $\text{Prob}\left(A_i + T_g > 0 \middle| T_g\right)$ is increasing in T_g. □

---

self-selection can provide additional information in identifying social interactions in linear models.

This proposition is useful as it indicates how the presence of endogenous social interactions may be inferred if $T_g > T_{g'}$ yet $E\left(m_g \middle| T_g\right) < E\left(m_{g'} \middle| T_{g'}\right)$. As discussed in the proof of Proposition 4, this can only occur, under the specification we have assumed, if group g has coordinated on an equilibrium expected average choice level other than the largest of the possible equilibria associated with the it while group g' has coordinated on an equilibrium other than the lowest possible expected average choice level among those it could have attained. The existence of multiple equilibria immediately implies $J > 0$.

The use of assortative matching to facilitate identification may be extended to panel data. To do this, modify (38) and (39) so that $A_{i,t} = k + cX_{i,t} + \varepsilon_{i,t}$ and $T_{g,t} = dY_g + eY_{g,t}$ and assumption *S.1* is modified to

S.2. For any two groups g and g', if $T_{g,t} > T_{g',t}$, then $F_{A_i|T_{g,t}}$ first order stochastically dominates $F_{A_i|T_{g',t}}$.

This modification allows one to identify implications for $m_{g,t}$ as time varies. For example, this Corollary to Proposition 9 is immediate.

Corollary 2. Equality of average outcomes with equal observable contextual effects.

If $J = 0$ or $J > 0$ but sufficiently small that $m_{g,t}$ is unique, then $Y_{g,t} = Y_{g,t'}$ implies $m_{g,t} = m_{g,t'}$ under assumptions $A.2,\ A.3,\ A.4,\ A.5'',\ A.6$ and *S.2*.

This corollary is useful because it indicates how evidence of endogenous social interactions may be adduced from temporal changes in $m_{g,t}$ when $Y_{g,t}$ does not vary. As such this result complements the identification strategy of Hoxby which relies on fluctuations in $Y_{g,t}$.

6. Conclusions

This paper provides a set of conditions under which identification holds for the binary choice model with social interactions. Relative to previous work, the analysis establishes identification conditions without assuming that the distribution for random payoff terms is logistic, as is done in Brock and Durlauf (2001a,b). In addition, some partial identification results are developed for models with unobserved group level variables.

With respect to further work, we mention two difficulties with the analysis found here. The analysis in the paper is heavily dependent on the fact that the distribution of random payoff terms may be identified based on within-group observations. For contexts such as classrooms, in which the number of group members is small, accurate finite sample approximations to this distribution may prove to be problematic. This suggests the importance of developing ways to combine information across groups to improve the accuracy of the estimation of this distribution. Second, the paper assumes that the groups within which interactions occur are themselves known a priori. As Manski (2000) has emphasized, this is a very strong assumption and may not be plausible in many cases. The absence of a coincidence between measured social groups and true social groups will induce complicated patterns of interdependences in errors across individuals as well as make it difficult to assess counterfactuals such as the effects of changes in the compositions of measured groups; examples of such counterfactuals include changes in the location of public housing or changes in educational policies regarding the tracking of students. Hence, the extension of the measuring of social interaction effects to policy evaluation will require assessing potential model uncertainty of this type. Some of this work is currently under way.

References

Aizer, A. and J. Currie, (2002), "Networks or Neighborhood? Correlations in the Use of Publicly-Funded Maternity Care in California," *NBER Working Paper no. 9209.*

Akerlof, G., (1997), "Social Distance and Economic Decisions," *Econometrica*, 65, 5, 1005-1027.

Angrist, J. and K. Lang, (2002), "How Important Are Classroom Effects? Evidence From Boston's Metco Program," *National Bureau of Economic Research Working Paper no. 9263*.

Anselin, L., (2003), "Spatial Externalities, Spatial Multipliers and Spatial Econometrics," *International Regional Science Review*, 26, 2, 153-166.

Barrett, G. and S. Donald, (2003), "Consistent Tests for Stochastic Dominance," *Econometrica*, 71, 1, 71-104.

Bayer, P. and C. Timmins, (2002), "Identifying Social Interactions in Endogenous Sorting Models," mimeo, Department of Economics, Yale University.

Becker, G., (1973), "A Theory of Marriage: Part I, "*Journal of Political Economy*, 81, 4, 813-846.

Bénabou, R., (1993), "Workings of a City: Location, Education, and Production," *Quarterly Journal of Economics*, 108, 3, 619-652.

Bénabou, R., (1996a), "Equity and Efficiency in Human Capital Investment: The Local Connection," *Review of Economic Studies*, 63, 2, 237-264.

Bertrand, M., E. Luttner, and S. Mullainathan, (2000), "Network Effects and Welfare Cultures," *Quarterly Journal of Economics,* 115, 1019-1055.

Bisin, A., A. Moro, and G. Topa, (2002), "The Empirical Content of Models with Multiple Equilibria," mimeo, New York University.

Boozer, M. and S. Cacciola, (2001), "Inside the Black Box of Project Star: Estimation of Peer Effects Using Experimental Data," *Yale University Economic Growth Center Discussion Paper no. 832.*

Brock, W., (1993), "Pathways to Randomness in the Economy: Emergent Nonlinearity and Chaos in Economics and Finance," *Estudios Economicos*, 8, 3-55.

Brock, W., W. D. Dechert, B. LeBaron and J. Scheinkman, (1996), "A Test for Independence Based on Upon the Correlation Dimension," *Econometric Reviews*, 15, 3, 197-235.

Brock, W. and S. Durlauf, (2001a) "Discrete Choice with Social Interactions," *Review of Economic Studies*, 68, 235-260.

Brock, W. and S. Durlauf, (2001b), "Interactions-Based Models," in J. Heckman and E. Leamer, eds., *Handbook of Econometrics 5*, Amsterdam: North-Holland.

Brock, W. and S. Durlauf, (2003), "A Multinomial Choice Model with Social Interactions," *The Economy as an Evolving Complex System III*, L. Blume and S. Durlauf, eds., Oxford University Press, forthcoming.

Cameron, S. and J. Heckman, (1998), "Life Cycle Schooling and Dynamic Selection Bias: Models and Evidence for Five Cohorts of American Males," *Journal of Political Economy*, 106, 2, 262-333.

Chamberlain, G., (1984), "Panel Data," in *Handbook of Econometrics, volume 2*, Z. Griliches and D. McFadden eds., Amsterdam: North-Holland, 1247-1318.

Conley, T. and G. Topa, (2002), "Socio-Economic Distance and Spatial Patterns in Unemployment," *Journal of Applied Econometrics*, 17, 4, 303-327.

Cooper, R., (1999), *Coordination Games*, New York: Cambridge University Press.

Crane, J., (1991), "The Epidemic Theory of Ghettos and Neighborhood Effects on Dropping Out and Teenage Childbearing," *American Journal of Sociology*, 96, 1226-1259.

Durlauf S., (1996a), "A Theory of Persistent Income Inequality," *Journal of Economic Growth*, 1, 1, 75-93.

Durlauf, S., (1996b), "Neighborhood Feedbacks, Endogenous Stratification, and Income Inequality," in *Dynamic Disequilibrium Modelling*, W. Barnett, G. Gandolfo, and C. Hillinger, eds., Cambridge: Cambridge University Press.

Durlauf, S., (2003), "Neighborhood Effects," *Handbook of Regional and Urban Economics, volume 4*, J. V. Henderson and J.-F. Thisse, eds., Amsterdam: North-Holland, forthcoming.

Evans, W., W. Oates, and R. Schwab, (1992), "Measuring Peer Group Effects: A Study of Teenage Behavior," *Journal of Political Economy*, 100, 5, 966-991.

Glaeser, E., B. Sacerdote, and J. Scheinkman, (1996), "Crime and Social Interactions," *Quarterly Journal of Economics*, 111, 2, 507-548.

Graham, B. and J. Hahn, (2003), "Identification and Estimation of Linear-in-Means Models," mimeo, Department of Economics, Harvard University.

Granger, C. and T. Terasvirta, (1996), *Modelling Nonlinear Economic Relationships*, New York: Oxford University Press.

Hanushek, E., J. Kain, J. Markman, and S. Rivkin, (2003), "Does the Ability of Peers Affect Student Achievement?," *Journal of Applied Econometrics*, 18, 527-544.

Hoff, K., and A. Sen, (2002), "Home-Ownership, Community Interactions, and Segregation," mimeo, World Bank.

Hoxby, C., (2000a), "The Effects of Classroom Size on Student Achievement: New Evidence from Population Variation," *Quarterly Journal of Economics*, 1239-1285.

Hoxby, C., (2000b), "Peer Effects in the Classroom: Learning from Gender and Race Variation," *National Bureau of Economic Research Working Paper no. 7867.*

Ioannides, Y. and J. Zabel, (2003a), "Neighborhood Effects and Housing Demand," *Journal of Applied Econometrics*, forthcoming.

Ioannides, Y. and J. Zabel, (2003b), "Interactions, Neighborhood Selection, and Housing Demand," mimeo, Department of Economics, Tufts University.

Krauth, B., (2002), "Simulation-Based Estimation of Peer Effects," mimeo, Department of Economics, Simon Fraser University.

Lindsay, B., (1995), *Mixture Models: Theory, Geometry and Applications*, Hawward: Institute for Mathematical Statistics.

Loury, G. (1977), "A Dynamic Theory of Racial Income Differences," in *Women, Minorities, and Employment Discrimination*, P. Wallace and A. Lamond, eds., Lexington, Massachusetts: Lexington Books.

Manski, C., (1988), "Identification of Binary Choice Models," *Journal of the American Statistical Association*, 83, 403, 729-738.

Manski, C., (1990), "Nonparametric Bounds on Treatment Effects," *American Economic Review*, 80, 319-323.

Manski, C., (1993), "Identification of Endogenous Social Effects: The Reflection Problem," *Review of Economic Studies*, 60, 531-542.

Manski, C. (1997), "Monotone Treatment Response," *Econometrica*, 65, 1311-1334.

Manski, C., (2000), "Economic Analysis of Social Interactions," *Journal of Economic Perspectives*, 14, 3, 115-136.

Manski, C. and J. Pepper, (2000), "Monotone Instrumental Variables: With an Application to Returns to Schooling," *Econometrica*, 70, 997-1010.

Manski, C., (2003), *Partial Identification of Probability Distributions*, New York: Springer-Verlag.

Nechyba, T., (2001), “Social Approval, Values and AFDC: A Re-Examination of the Illegitimacy Debate,” *Journal of Political Economy*, 109, 3, 637-666.

Rosenbaum, P. and D. Rubin, (1983), “The Central Role of the Propensity Score in Observational Studies for Causal Effects,” *Biometrika*, 70, 41-55.

Sampson, R., Morenoff, J., and Earls, F., (1999), “Beyond Social Capital: Collective Efficacy for Children,” *American Sociological Review*, 64, 633-660.

Sampson, R., Raudenbush, S., and Earls, F., (1997), “Neighborhoods and Violent Crime: A Study of Collective Efficacy,” *Science*, 277, 918-924.

Sattinger, M., (1993), “Assignment Models of the Distribution of Earnings, *Journal of Economic Literature*, June 1993, 31, 2, 831-880.

Schelling, T., (1971), “Dynamic Models of Segregation,” *Journal of Mathematical Sociology*, 1, 2, 143-186.

Sirakaya, S. (2002), “Recidivism and Social Interactions,” mimeo, University of Wisconsin.

South, S. and E. Baumer, (2000), “Deciphering Community and Race Effects on Adolescent Premarital Childbearing,” *Social Forces*, 78, 4, 1379-1408.

Topa, G., (2001), “Social Interactions, Local Spillovers, and Unemployment, *Review of Economic Studies*, 68, 2, 261-295.