TECHNICAL WORKING PAPER SERIES

MULTINOMIAL CHOICE WITH SOCIAL INTERACTIONS

William A. Brock
Steven N. Durlauf

Technical Working Paper 288
http://www.nber.org/papers/T0288

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
February 2003

Multinomial Choice with Social Interactions
William A. Brock and Steven N. Durlauf
NBER Technical Working Paper No. 288
February 2003
JEL No. C35, C72

ABSTRACT

This paper develops a model of individual decisionmaking in the presence of social interactions when the number of available choices is finite. We show how a multinomial logit model framework may be used to model such decisions in a way that permits a tight integration of theory and econometrics. Conditions are given under which aggregate choice behavior in a population exhibits multiple self-consistent equilibria. An econometric version of the model is shown to be identified under relatively weka conditions. That analysis is extended to allow for general error distributions and some preliminary ways to account for the endogeneity of group memberships are developed.

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

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

Multinomial Choice with Social Interactions

William A. Brock and Steven N. Durlauf*

University of Wisconsin at Madison

January 27, 2003

For if anyone, no matter who, were given the opportunity of choosing amongst all the nations of the world the beliefs he thought best, he would inevitably, after careful consideration of their relative merits, choose those of his own country. Everyone without exception believes his own native customs, and the religion he was brought up in, to be the best…There is abundant evidence that this is the universal feeling about the ancient customs of one's country. One might recall…an account told of Darius. When he was king of Persia, he summoned the Greeks who happened to be at this court and asked them what they would take to eat the dead bodies of their fathers. They replied that they would not do it for any money in the world. Later, in the presence of the Greeks…he asked some Indians…who do in fact eat their parents' dead bodies, what they would take to burn them. They uttered a cry of horror and forbade him to mention such a dreadful thing. One can see by this what custom can do and Pindar, in my opinion, was right when he called it 'king of all.'

Herodotus, *The Histories* (3.38)[1]

This paper provides a model of individual decisionmaking in the presence of social interactions. By social interactions, we refer to interdependencies between individual decisions and the decisions and characteristics of others within a common group. In virtually any economic model, the decisions of one agent will be influenced by the behaviors and characteristics of others; what distinguishes the perspective we adopt is that the interdependences we study directly link individuals. By way of contrast, agents in a market are influenced by a common

---

*Department of Economics, University of Wisconsin, 1180 Observatory Drive, Madison, WI 53706-1393. We thank the John D. and Catherine T. MacArthur Foundation, National Science Foundation, University of Wisconsin Graduate School and Vilas Trust for financial support. Artur Minkin, Eldar Nigmatullin and Chih Ming Tan have provided outstanding research assistance. This paper is written in honor of Kenneth Arrow.

[1]Taken from Herodotus, *The Histories*, A. de Selincourt, trans., New York: Penguin, 1996.

price that reflects the participation of each individual in the market. The sorts of phenomena we study are different as these describe ways to formalize ideas such as peer group influences, whereby behaviors of one agent alter the preferences of others and are not mediated by how individuals affect prices in a market equilibrium. As such, social interactions constitute an example of the type of externalities described in Arrow and Hahn (1971, chapter 6, section 2).

Within economics, there has developed an increasing recognition that social interactions may play a major role in explaining a range of individual behaviors. In many respects, the new literature on social interactions addresses a famous criticism of economics made in Granovetter (1985),

> "Classical and neoclassical economics operates…with an atomized and *under*socialized conception of human action…The theoretical arguments disallow by hypothesis any impact of social structure and social relations…" (pg. 55)

In fact, one of the appealing aspects of the new literature on social interactions is that it has facilitated the introduction of sociological concepts and perspectives into economic modeling. In turn, the economics literature on social interactions has shown how these ideas may be formalized and extended using the formal rigor of economic theory. More important, the approach we take represents a first step in an integration of individual-based and social-based explanations, a perspective whose importance is well described in Arrow (1994)

> "It is clear that the individualist perspective does play an essential role in understanding social phenomena. Particularly striking is the *emergent* nature of social phenomena, which may be very far from the motives of the individual interactions. " (pg. 3)

As such, we regard the social interactions literature as a successful example of how social science benefits from the breaking down of disciplinary barriers.

There is now a large body of theoretical and empirical studies of social interactions. In terms of theory, two main approaches have been taken. One strand of the social interactions literature has focused on the implications of social interactions in predetermined groups. Akerlof (1997) and Brock and Durlauf (1999,2001a,b), for example, consider the role of the interactions structure within a group on group-level outcomes. Models such as Loury (1977) and Lundberg and Startz (1998) focus on the effects of social interactions within ethnic groups with specific attention to how differences in initial conditions have long run effects. In contrast, work by Bénabou (1993,1996), Durlauf (1996a,b) and Hoff and Sen (2000) has primarily focused on the implications of social interactions for group formation, specifically in the context of residential neighborhoods. In these models, children are influenced by the neighborhoods in which they grow up through factors such as the local tax base, the types of role models that are present and via peer group influences. Models of this type can produce poverty traps as poverty among parents is transmitted to children when children are consigned to neighborhoods whose interactions adversely affect their subsequent economic status; poverty among parents, because of its affect on children's neighborhoods, thus transmits economic status across generations. One limitation of the existing theoretical models of social interactions is the relatively weak connections between these two approaches.[2]

The empirical literature on social interactions has been dominated by attention to the effects of residential neighborhoods. A wide range of analyses have produced regression evidence that links individual outcomes with neighborhood (i.e. groups defined by geographic proximity) characteristics; examples include Aizer and Currie (2002), Brooks-Gunn et al (1993), Corcoran et al (1992), Brewster (1994), Datcher (1982), Ginther, Haveman and Wolfe (2000), Ioannides and Zabel, (2002a,b), Nigmatullin (2002), Plotnik and Hoffman (1999), Sirakaya (2002), South and Baumer (2000), and South and Crowder (1999). Alternative strategies for uncovering neighborhood effects using aggregated data have been developed by Glaeser, Sacerdote, and Scheinkman (1996) and Topa (2001). In addition, there is now a literature that moves beyond the assumption that geographic proximity determines interactions and attempts to identify which sorts of groups in fact produce social interactions. Aizer and Currie (2002) and Conley and Topa (2002) are interesting empirical analyses that compare alternative groups (e.g. groups defined by geographic proximity versus common ethnicity) in terms of the social interactions with which they are associated.

Empirical work on neighborhood effects has been buttressed by two recent developments. The first is the use of "quasi-experimental" data produced by government interventions that alter the neighborhood choices of certain families. Examples of such programs include the Gautreaux

---

[2] See Becker and Murphy (2000) for a synthesis of various theoretical strands of the social interactions literature as well as for a valuable analysis of links between the two strands we have described.

program (Rosenbaum (1995) and Rosenbaum and Popkin (1991)) and the Moving to Opportunity Demonstration (Katz, Kling, and Liebman (2001), Ludwig, Duncan and Hirshfeld (2001)). Each of these programs is interesting because each constitutes a government intervention in which a set of poor families are given incentives to move to more affluent neighborhoods, thereby permitting comparison with similar families who did not receive such incentives. These studies generally find improved outcomes, especially for children, among families that move.

The second is the development of a detailed data set that measures attitudes and beliefs across neighborhoods, called the Project on Human Development in Chicago Neighborhoods. As illustrated in Sampson, Morenoff, and Earls (1999) and Sampson, Raudenbush and Earls (1997), this data can illuminate some of the structural relationships that underlie the correlations that are found in other studies between neighborhood attributes and individual outcomes. For example, these studies find that "collective efficacy," by which they mean the willingness of neighborhood members to provide public goods that contribute to social order (one example is monitoring the children of neighbors) appears strongly associated with lower crime rates.

A major weakness of the social interactions literature as it is currently constituted is the absence of strong connections between theory and empirics. By this, we mean that there has been little effort to employ theoretical models of social interactions in structural estimation. Our previous work on social interactions, Brock and Durlauf (2001a,b) has attempted to address this limitation by developing models of binary choice that are directly econometrically implementable. This paper extends that work to account for multinomial choice. This generalization leads to a number of new methodological insights as well as allows for the application of theoretical models of social interactions to a broader range of phenomena than was previously possible.

Section I outlines a basic choice model with social interactions. Section II analyzes a version of this general framework in which individual choices follow a multinomial logit structure. Section III considers the econometric implementation of the multinomial logit model we have developed with specific attention to identification problems that may arise when social interactions are present. Section IV discusses how to extend our basic framework to account for alternative error specifications. We show that the basic theoretical and econometric features of the multinomial logit model apply quite generally. Section V considers how to integrate decisions on behaviors with decisions on group memberships. Section VI contains summary and conclusions. A Technical Appendix contains all proofs.

I. Modeling social interactions

We consider I individuals who are members of a common group g. Our objective is to probabilistically describe the individual choices of each i, ω_i (a choice that is taken from the elements of some set of possible behaviors Ω_i) and thereby characterize the vector of choices of all members of the group, ω.

From the perspective of theoretical modeling, it is useful to distinguish between three sorts of influences on individual choices. These influences have different implications for how one models the choice problem. These components are

h_i, a vector of deterministic (to the modeler) individual-specific characteristics associated with individual i,

ε_i, a vector of random individual-specific characteristics associated with i,

and

$\mu_i^e(\omega)$, the subjective beliefs individual i possesses about behaviors in the group, expressed as a probability measure over those behaviors.

Each of these components will be treated as a distinct argument in the payoff function that determines individual choices. As we shall see, each plays a distinct role in the analysis.

The deterministic and random individual-specific characteristics capture the "standard" determinants of individual choices that one finds in economic models. Hence, if ω_i represents whether one is deciding between college and employment, or military enlistment, h_i may include variables such as educational attainment of one's parents, or the quality of one's high school; the

ε_i vector may include unobservable variables such as "true" intellectual ability or tastes. The distinction between deterministic and random characteristics will prove to play an important role in both the theoretical and the econometric analysis of the model.

From the perspective of individual decisionmaking, what distinguishes our framework from the standard discrete choice approach is the possibility that individual choices are affected by $\mu_i^e(\omega)$, the beliefs an individual has about others. This interdependence in fact lies at the heart of the new literature on social interactions. In modeling social interactions across individuals, we assume that these interactions are mediated by beliefs, i.e. individual i is influenced by what he thinks others are doing, not by their actual behavior per se. This assumption provides a great deal of analytical convenience. Its appropriateness will depend on the context under study, and in particular the size of the group in which interactions occur. So, while the assumption seems relatively natural when one is interested in social interactions at an ethnic group level, it is clearly problematic in describing interactions between a pair of best friends.

Individual choices ω_i are characterized as representing the maximization of some payoff function V,

$$\omega_i = \arg\max_{\lambda \in \Omega_i} V\left(\lambda, h_i, \mu_i^e(\omega), \varepsilon_i\right) \tag{1}$$

Thus, we treat the decision problem facing an individual as a function of preferences (embodied in the specification of V), constraints (embodied in the specification of Ω_i) and beliefs (embodied in the specification of $\mu_i^e(\omega)$). As such, our analysis is based on completely standard microeconomic reasoning to describe individual decisions.

This basic choice model is closed by imposing self-consistency between subjective beliefs $\mu_i^e(\omega)$ and the objective conditional probabilities $\mu(\omega \mid F_i)$, where F_i denotes the information available to agent i. We assume that agents know the deterministic characteristics of others as well as themselves and also understand the structure of the individual choice problems that are being solved. This means that subjective beliefs must obey

$$\mu_i^e(\omega) = \mu\left(\omega \middle| h_j, \mu_j^e(\omega)\ \forall j\right) \tag{2}$$

where the right hand of this equation is the objective conditional probability measure generated by the model; self-consistency is equivalent to rational expectations in the usual sense. From the perspective of modeling individual behaviors, it is typically assumed that agents do not account for the effect of their choices on the decisions of others via expectations formation. In this sense, this framework embodies an expectations-based version of a Nash equilibrium.

This general structure illustrates how social interactions models preserve the individual choice-based logic of microeconomics. Their novelty lies in the interdependences in choices that are induced by including $\mu_i^e(\omega)$ as an argument in individual payoffs and imposing self-consistency. From the perspective of economic theory, the interesting properties of these models emerge as a result of this interdependence.

II. A multinomial logit approach to social interactions

i. basic setup

In order to understand the implications of social interactions for the equilibrium distribution of choices within a population, it is necessary to specialize this general behavioral description. We do this in three steps.

First, we assume that each agent faces a common choice set with L discrete possibilities, i.e. $\Omega_i = \{0, 1, \ldots, L-1\}$. The common choice set assumption is without loss of generality, since if agents face different choice sets, one can always assume their union is the common set and then specify that certain choices have payoff of $-\infty$ for certain agents.

Second, we assume that each choice l produces utility for i according to:

$$V_{i,l} = h_{i,l} + Jp_{i,l}^e + \varepsilon_{i,l} \tag{3}$$

Following the notation of the previous section, $h_{i,l}$ denotes the deterministic private utility agent i receives from the choice, $Jp_{i,l}^{e}$ denotes the social utility from the choice, and $\varepsilon_{i,l}$ denotes random private utility from the choice. The social utility term contains both a measure of the strength of social utility, J, and the agent's subjective expectation of the percentage of agents in the neighborhood who make the same choice $p_{i,l}^{e}$. This is the natural generalization of the conformity effect model developed in Brock and Durlauf (2001a,b) and is also employed in Bayer and Timmins (2001).

Third, we assume that the errors $\varepsilon_{i,l}$ are independent across i and are doubly exponentially distributed with index parameter β, i.e.

$$\mu\left(\varepsilon_{i,l} \leq \varsigma\right) = \exp\left(\exp\left(-\beta\varsigma + \gamma\right)\right) \tag{4}$$

where γ is Euler's constant. This assumption is of course standard in the discrete choice literature and is the basis of multinomial logit specifications; see Anderson, de Palma and Thisse (1992) for discussion of the substantive behavioral restrictions that this specification imposes. The parameter β measures the dispersion in the random utility terms; higher β implies lower dispersion.

These assumptions may be combined to produce a full description of the choice probabilities for each individual, $\mu\left(\omega_i = l \middle| h_{i,j}, p_{i,j}^{e}\ \forall j\right)$. The probability that agent i chooses l equals the probability that the payoff associated with this choice is maximal among all payoffs available to the agent, i.e.

$$\mu\left(\omega_i = l \middle| h_{i,j}, p_{i,j}^{e}\ \forall j\right) = \mu\left(\arg\max_{j\in\{0\ldots L-1\}} h_{i,j} + Jp_{i,j}^{e} + \varepsilon_{i,j} = l \middle| h_{i,j}, p_{i,j}^{e}\ \forall j\right) \tag{5}$$

This is a standard calculation under the double exponential assumption for the random payoff terms and leads to the canonical multinomial logit probability structure (cf. Anderson, dePalma and Thisse (1992, section 2.6)):

$$\mu\left(\omega_i = l \middle| h_{i,j}, p^e_{i,j}\ \forall j\right) = \frac{\exp\left(\beta h_{i,l} + \beta J p^e_{i,l}\right)}{\sum_{j=0}^{L-1}\exp\left(\beta h_{i,j} + \beta J p^e_{i,j}\right)} \tag{6}$$

When there are only two choices, this is the model studied in Brock and Durlauf (2001a,b). Since the random payoff terms are independent across agents, the joint choice probabilities may be written as

$$\mu\left(\omega_1 = l_1, \ldots, \omega_I = l_I \middle| h_{i,j}, p^e_{i,j}\ \forall i, j\right) = \prod_i \frac{\exp\left(\beta h_{i,l_i} + \beta J p^e_{i,l_i}\right)}{\sum_{j=0}^{L-1}\exp\left(\beta h_{i,j} + \beta J p^e_{i,j}\right)} \tag{7}$$

Finally, we characterize self-consistency as defined by eq. (2). We assume that the information set of each agent includes both the empirical distribution of deterministic payoff terms across choices and agents as well as the probability distribution from which random utility terms are drawn. We also assume that the number of agents is sufficiently large that each agent ignores the effect of his own choice on the average.[3] As self-consistent beliefs imply that the subjective choice probabilities p^e_l equal the objective expected values of the percentage of agents in the group who choose l, p_l, the structure of the model implies that

$$p^e_{i,l} = p_l = \int \frac{\exp\left(\beta h_{i,l} + \beta J p_l\right)}{\sum_{j=0}^{L-1}\exp\left(\beta h_{i,j} + \beta J p_j\right)} dF_h \tag{8}$$

---

[3] In cases where the number of agents is small, it is perhaps more natural to express an individual's payoff as depending on the actual choices of others. There are subtle issues that need to be dealt with in the small numbers case as this essentially means that agents know the $\varepsilon_{i,l}$'s for others in the group; see Kooreman and Soetevent (2002). Nevertheless, a small group approach is closely related to our framework.

where F_h is the empirical probability distribution for the vector of deterministic terms $h_{i,l}$. It is straightforward to verify that under the Brouwer fixed point theorem, at least one such fixed point exists, so this model always has at least one equilibrium set of self-consistent aggregate choice probabilities.

ii. properties

To understand the properties of this model, it is useful to focus on the special case where $h_{i,l} = 0 \ \forall i,l$. For this special case, the choice probabilities (and hence the expected distribution of choices within a neighborhood) are completely determined by the compound parameter βJ. An important question is whether and how the presence of interdependencies produces multiple equilibria for the choice probabilities in a neighborhood. In order to develop some intuition as to why the number of equilibria is connected to the magnitude of βJ, it is helpful to consider two extreme cases for the compound parameter, namely $\beta J = 0$ and $\beta J = \infty$. We consider each case in turn.

For the case $\beta J = 0$, one can immediately verify that there exists a unique equilibrium for the aggregate choice probabilities such that $p_l = \frac{1}{L} \ \forall l$. This follows from the fact that under the assumption that $h_{i,l} = 0 \ \forall i,l$, all individual heterogeneity in choices come from the realizations of $\varepsilon_{i,l}$, a process whose elements are independent and identically distributed across choices and individuals. Since all agents are ex ante identical, the aggregate choice probabilities must be equal.

The case $\beta J = \infty$ is more complicated. The set of aggregate choice probabilities $p_l = \frac{1}{L}$ is also an equilibrium if $\beta J = \infty$ since conditional on these probabilities, the symmetries in payoffs associated with each choice that led to this equilibrium when $\beta J = 0$ are preserved as there is no difference in social utility across choices. However, this is not the only equilibrium. To see why this is so, observe that for any pair of choices l and l' for which the aggregate choice probabilities are nonzero, it must be the case that

$$\frac{p_l}{p_{l'}} = \frac{\exp(\beta J p_l)}{\exp(\beta J p_{l'})} \tag{9}$$

for any βJ. This follows from the fact that each agent is ex ante identical. Thus, it is immediate that any set of equilibrium probabilities that are bounded away from 0 will become equal as $\beta J \Rightarrow \infty$. This condition is necessary as well as sufficient, so any configuration such that $p_l = \frac{1}{b}$ for some subset of b choices and $p_l = 0$ for the other $L-b$ choices is an equilibrium. Hence, for the case where $J = \infty$, there exist $\sum_{b=1}^{L} \binom{L}{b} = 2^L - 1$ different equilibrium probability configurations. Recalling that β indexes the density of random utility and J measures the strength of interdependence between decisions, this case, when contrasted with $\beta J = 0$ illustrates why the strength of these interdependences and the degree of heterogeneity in random utility interact to determine the number of equilibria.

These extreme cases may be refined to produce a more precise characterization of the relationship between the number of equilibria and the value of βJ. In general, this relationship is highly complex as it is necessary to account for the distribution of $h_{i,l}$ across i and l within a given group in order to develop precise statements as to how the model parameters determine the number of equilibria. Theorem 1 characterizes how the magnitude of βJ determines the number of equilibria in this case.

Theorem 1. Multiple equilibria in the multinomial logit model with social interactions

Suppose that individual choices are characterized by eq. (6) with self-consistent beliefs, i.e., that beliefs are consistent with eq. (8). Assume that $h_{i,l} = k \ \forall i,l$. Then there will exist at least three self-consistent choice probabilities if $\frac{\beta J}{L} > 1$.

When $L = 2$, this theorem reduces to the characterization of multiple equilibria with binary choices in Brock and Durlauf (2001a).[4]

In general, it is difficult to extend Theorem 1 to account for cases where $h_{i,l}$ is nonconstant. The reason for this is that the equilibrium aggregate choice probabilities induced by the interaction of private incentives and social incentives will in general depend on the complete distribution of $h_{i,l}$ across choices and individuals. There are some special cases where one can establish precise results. For example, suppose that $h_{il} - h_{i0} = g \ \forall l \in \{1 ... L-1\}$. In this case, the private deterministic utility of choice 0 differs from that of the other choices. In this case, the proof of Proposition ii in Brock and Durlauf (2001b) implies that if $\frac{\beta J}{L} > 1$ there exists a threshold T for g such that if $0 < g < T$ there are multiple equilibria whereas if $g > T$ the equilibrium is unique.

There is a common basic intuition for Theorem 1 and similar results in Brock and Durlauf (2001a,b) that relate the number of choice equilibria to the interplay between the strength of social utility, J, the levels of deterministic private payoffs, $h_{i,l}$, and the parameter that indexes the degree of dispersion in random private utility, β. Multiple equilibria arise when the social utility effects on individual behavior can induce self-consistent bunching on a subset of choices. A positive J induces a tendency towards self-consistent bunching. Such a tendency is counteracted by the private utility components. One way in which this tendency towards self-consistent bunching may be countered is via the distribution of $h_{i,l}$; these private deterministic payoff components can, if sufficiently skewed, render the aggregate choice probabilities unique. A similar effect can occur via the realizations of the random payoff terms $\varepsilon_{i,l}$. With respect to Theorem 1, if the random utility components are sufficiently dispersed (i.e. β is small), then a

---

[4] Brock and Durlauf (2001a,b) use slightly different normalizations for the analysis of binary choice. Specifically, choices are indexed -1 and 1 and the social utility component is written as Jm where m is the expected value of the choices in the group. For this reason, the threshold theorem in Brock and Durlauf (2001a,b) is stated in terms of whether or not $\beta J > 1$ rather than $\frac{\beta J}{2} > 1$ as is done here.

sufficient percentage of agents will have draws of random utility such that their choices are dominated by one of the $\varepsilon_{i,l}$'s regardless of the strength of social utility, leaving too small a percentage of agents to engage in self-consistent bunching, as the social utility associated with the bunching depends on the percentage of agents that make the choice. Put differently, the presence of social utility effects, considered in isolation, do not identify what choices an individual makes, only that choices across individuals will be correlated. This induces a degree of freedom in the determination of what choices are actually made. (This is the same intuition for the presence of multiple equilbria in various coordination failure models; see Cooper (1999) for a survey.) The individual-specific deterministic and random terms, considered in isolation, do produce unique choices for the population. The interplay between the strength of these influences determines the number of equilibrium choice configurations.

An interesting feature of Theorem 1 is the fact that the condition for multiple equilibria depends on the number of choices. As such, the theorem explains simulation evidence in Bayer and Timmins (2001) which indicated that multiple equilibria seem less likely in models when more choices are involved. This theorem makes their findings precise and provides some insight as to why they occur. Intuitively, the reason that the number of choices raises the threshold value of βJ necessary for multiple equilibria is the assumption that the random utility terms are independent. This independence means that the percentage of individuals in a population whose behavior is determined by their random utility (because of an extremely large draw for one of the choices relative to the others) increases in the number of choices, leaving a smaller percentage of the population susceptible to self-consistent bunching due to the influence of Jp_l. Higher percentages of agents whose behavior is determined by the random utility draws will reduce the potential for social utility to produce multiple self-consistent equilibria.

iii. cooperative equilibria

In this section, we consider the formulation of a cooperative analog to the noncooperative model we have studied in Section II.ii. The welfare properties of the noncooperative equilibria are best understood when explicitly contrasted with the equilibria that would occur under some sort of cooperation. One way to do this is to develop an analogous social planner's problem for the set of choices under consideration. Such an approach requires the use of relatively sophisticated models and techniques from the statistical mechanics literature. Following ideas originally developed in Brock (1993) and subsequently elaborated in the present context in Brock and Durlauf (2001a), Brock and Durlauf (2002) proposed a way of formulating the behavior of a particular benign social planner (i.e. one whose objective function includes the sum of the deterministic payoff components of the individual agents) in such a way that the social planner's choice of configuration ω is given by a probability measure $\mu(\omega)$ of the form

$$\mu(\omega) = Z_I^{-1} \exp\left(\sum_{i=1}^{I} \sum_{l=0}^{L-1} \left(\beta h_{i,l} 1(\omega_i = l) \right) + I\beta J \sum_{l=0}^{L-1} \hat{p}_l^2 \right) \tag{10}$$

In this expression, Z_I is a normalization and the $\hat{p}_l$'s are the empirical percentages of choices in the group, i.e.

$$\hat{p}_l = I^{-1} \sum_{i=1}^{I} 1(\omega_i = l) \tag{11}$$

where $1(\omega_i = l)$ denotes the indicator function for the choice of l by agent i.

In comparison with the probability measure that characterizes choices for noncooperative equilibrium (eq. (7)), the important difference is that the social planner's problem uses empirical probabilities in modeling the interdependence of individual choices whereas the noncooperative equilibrium is based on population probabilities (i.e. the agents' rational expectations concerning the choices of others.) This difference is to be expected since a planner will account for how the choices of one actor affect others in ways that are ruled out in the noncooperative case. This feature makes the probability structure much more difficult to analyze. For example, the joint probability measure for the planner does not factor into a product of marginal probabilities (each representing one individual's choice) as it does in the noncooperative case. This means that there is a direct channel by which each agent's choice becomes correlated with the choices of others. The nature of this direct dependence as $I \Rightarrow \infty$ plays a key role in determining the aggregate behavior of the population.

We conjecture that as $I \Rightarrow \infty$, $\hat{p}_l \Rightarrow_w p_l^* \ \forall l$, ($w$ denotes weak convergence), the vector p^* with typical element p_l^* solves

$$p^* = \arg\max_p \lim_{I \Rightarrow \infty} I^{-1} \ln Z_I \tag{12}$$

The Brock and Durlauf (2002) assertion (their equation (12)) that (12) holds is incorrect as stated because the sufficient conditions for (12) to hold are left out. For example, when $L = 2$ and $h_{i,l} = 0$, if $\frac{\beta J}{2} > 1$ then there are two global maxima to (12) which means that the limit of the sample mean $\hat{p}$ is a mixture with a two point support; Ellis (1986 p. 100) provides a complete analysis of the binary case when $h_{i,l} = h$. For the binary case with random $h_{i,l}$, results by Amaro de Matos and Perez (1991) may be adapted to locate sufficient conditions for a result such as (12) to hold. In fact, for the binary choice case, these results suggest that the usual central limit theory for suitably normalized sums such as $\sum_i (\omega_i - E(\omega_i))$ needs to be modified. Although the usual central limit theorem results hold as long as 1) the value of βJ does not equal the critical value around which multiple equilibria emerge and 2) $h_{i,l}$ is constant (Ellis (1985, Theorem V.9.4)), the situation changes when the variance of $h_{i,l}$ is positive even when the global maximum of (10) is unique and various regularity conditions are imposed (Amaro de Matos and Perez (1991, Theorem 2.8, (a))). More precisely, the appropriately normalized sum of deviations around the mean will converge to a mixture of normals, not a normal distribution as occurs in standard cases. Further, small changes in the distribution of $h_{i.l}$ can lead to large differences in the global maximum of (12).

We do not know whether a result such as (12) holds for the multinomial case with general $h_{i,l}$. For the case where the global maximum to (10) is unique, there are a number of existing results that suggest that our conjecture is true. For example, Ellis and Wang (1990) analyze the model (10) where $h_{i,l} = 0 \quad \forall i, l$ and show there is a threshold J_T such that if $J < J_T$, then $\hat{p}_l$ converges weakly to $\frac{1}{L}$. We will pursue a full analysis of (12) in subsequent work.

III. Econometric implementation

i. Basic framework

An important feature of the theoretical framework is that it can also be used for econometric analysis.[5] The multinomial logit property for the individual choices allows one to construct a likelihood function for data taken for I individuals who are sampled across groups. Since a typical data set will contain observations on individuals in different groups, we generalize our notation so that $g(i)$ denotes the group of agent i; ω is now the vector of choices in a given cross-group sample of individuals. Finally, each individual within a group is modeled as possessing identical beliefs about the percentage of choices within the group, so that for choice l within group $g(i)$ each group member shares a common belief concerning the expected percentage of group members that are choosing l, $p^e_{g(i),l}$. Following Manski (1993), the dependence of individual behavior on $p^e_{g(i),l}$ is known as an endogenous effect, in order to highlight the notion that the (expected) choices of one agent influence the choices of another.

In empirical work on neighborhood effects the generic deterministic private incentive $h_{i,l}$ is usually assumed to depend on two types of observables: an *r*-dimension vector of individual characteristics X_i and an *s*-dimension vector of neighborhood characteristics $Y_{g(i)}$, also known as contextual effects. Manski (1993) provides the first formal discussion of this dichotomy, which is irrelevant to the development of the theory of social interactions but has important implications for econometric analysis. Operationally, it is standard to assume

---

[5] A range of econometric issues that arise for models of social interactions have been studied in Brock and Durlauf (2001a,b), Manski (1993), and Moffitt (2001). Brock and Durlauf (2001b) is the study that most closely focuses on issues concerning discrete choice models, also extending the analysis of identification to duration data.

$$h_{i,l} = k_l + c_l' X_i + d_l' Y_{g(i)}. \tag{13}$$

There is no necessary reason why the same elements of X_i and $Y_{g(i)}$ should affect the payoff of each choice; one can allow for this by setting particular elements of c_l and d_l to zero.

Under the assumption that $\varepsilon_{i,l}$ is independent of X_i and $Y_{g(i)}$ $\forall i,l$, the likelihood function for a collection of choices ω will equal

$$Z_I^{-1} \prod_i \left(\sum_l \exp\left(\beta k_l + \beta c_l' X_i + \beta d_l' Y_{g(i)} + \beta J_l p^e_{g(i),l} \right) 1\left(\omega_i = l\right) \right) \tag{14}$$

where Z_I is the normalization

$$Z_I = \prod_i \left(\sum_l \exp\left(\beta k_l + \beta c_l' X_i + \beta d_l' Y_{g(i)} + \beta J_l p^e_{g(i),l} \right) \right) \tag{15}$$

and beliefs are subject to a set of constraints on the subjective beliefs for members of each group $g(i)$,

$$p^e_{g(i),l} = \mathrm{E}\left(p_{g(i),l} \middle| F_{X_{g(i)}}, Y_{g(i)}, p^e_{g(i),l} \forall l \right) \tag{16}$$

where $F_{X_{g(i)}}$ is the empirical distribution of X_i within group $g(i)$ and expectations are formed on the basis of the probabilities defined by (15). This set of constraints imposes self-consistency in expected choice probabilities across groups and choices in the way that corresponds to the analysis in Section II.

As is standard for multinomial logit models, the complete set of model parameters is not identified. It is therefore necessary to impose some normalizations; we follow McFadden (1984, p. 1413) and impose the normalizations that $k_0 = 0,$ $c_0 = 0,$ $d_0 = 0$, $J_0 = 0$ and β=1.

ii. identification

As originally recognized and analyzed in Manski (1993) and further analyzed in Brock and Durlauf (2001b), Minkin (2002) and Moffitt (2001), there are possible identification problems in social interactions models due to the relationship between contextual effects $Y_{g(i)}$ and the equilibrium expected group choice probabilities $p_{g(i),l}$. Specifically, Manski (1993) shows how for a class of linear models of group effects, collinearity between particular contextual effects and endogenous effects (in our context, the $p_{g(i),l}$'s) that represent self-consistent beliefs about aspects of behaviors in the group can induce nonidentification. However, in contrast to the linear case, identification can hold for our model, as described in the following theorem.

Theorem 2. Identification of the multinomial choice model with neighborhood effects

Let the true data generating process be given by (14)-(16) with the normalization $k_0 = 0,\ c_0 = 0,\ d_0 = 0$, $J_0 = 0$ and β=1. Assume

i) the joint support of $X_i, Y_{g(i)}$ is not contained in a proper linear subspace of R^{r+s}

ii) the support of $Y_{g(i)}$ is not contained in a proper linear subspace of R^s ,

iii) no linear combination of elements of X_i and $Y_{g(i)}$ is constant,

iv) for each choice l, there exists at least one group g_l such that conditional on Y_{g_l} , X_i is not contained in a proper linear subspace of R^r ,

v) none of the elements of $Y_{g(i)}$ possesses bounded support,

vi) $p_{g(i),l}$ is not constant across neighborhoods,

vii) $\varepsilon_{i,l}$, the random utility terms for each individual, are independent of his associated X_i and $Y_{g(i)}$ and independent and identically distributed across choices and individuals.

Then the true set of model parameters $(k_1, c_1, d_1, J_1, \ldots, k_{L-1}, c_{L-1}, d_{L-1}, J_{L-1})$ is identified relative to any distinct alternative.

The proof of this theorem may be found in the appendix and is a generalization of a theorem on identification of neighborhood effects for binary choices found in Brock and Durlauf (2001a,b). The key to identification in this model is that, because models of discrete choice are inherently nonlinear in the various control variables (since choice probabilities are bounded), contextual effects and endogenous effects (in this case, the choice probabilities) cannot be linearly dependent. What the theorem in essence requires is three things. First, it is necessary that the data contain sufficient intraneighborhood variation within at least one neighborhood to ensure that k_l and c_l are identified $\forall l$. Second, there must be enough interneighborhood variation in $Y_{g(i)}$ to ensure that d_l and J_l are identified $\forall l$ because of the nonlinear relationship between contextual effects and endogenous effects. Third, there cannot be collinearity between the regressors contained in X_i and $Y_{g(i)}$, so that individual and contextual effects may be distinguished.

The conditions of the Theorem are sufficient, and clearly one could find weaker ones than those we have employed. An advantage of the conditions we have used is that they make clear what underlying properties are needed for identification and so should provide a guide to developing weaker conditions if needed in a particular context.

The identification theorem applies to more general models than that studied in Section III.ii. as the econometric model allows for a distinct J_l for each choice. This is appealing as one can easily imagine cases where the payoff from conforming to the behavior of others depends on the nature of the choice. For example, if one is choosing between a solitary and a group activity, one would intuitively expect the value of J_l to depend on the choice.

iii. extensions

Identification may also be established for the case where individual decisions depend on the expected percentages of individuals making each of the other choices as well as on the expected percentage of individuals making that choice. Formally, this means replacing (14) and (15) with

$$Z_I^{-1}\prod_i\left(\sum_l \exp\left(\beta k_l + \beta c_l' X_i + \beta d_l' Y_{g(i)} + \beta J_l' p^e_{g(i)}\right)1\left(\omega_i = l\right)\right) \tag{17}$$

and

$$Z_I = \prod_i\left(\sum_l \exp\left(\beta k_l + \beta c_l' X_i + \beta d_l' Y_{g(i)} + \beta J_l' p^e_{g(i)}\right)\right) \tag{18}$$

For (17) and (18), J_l is a vector $\left(J_{1,l} \dots J_{L-1,l}\right)$ and represents the weights, conditional on choice l that agent i assigns to the percentage of the population making each of the choices; $p^e_{g(i)}$ is the vector of expected choice percentages. Such a generalization is also appealing in various contexts. Suppose one is making a choice of religious affiliation in a population. It might be the case that the adherence to one affiliation is affected by the percentages of the population that adhere to certain other denominations For example, adherence to a particular affiliation that believes in creationism may be affected by the percentage of adherents to other Christian denominations that possess similar beliefs.[6] This generalization is also interesting because it allows for the possibility that there is negative social utility associated with particular cross-choice effects.[7] Hence, the expected percentage of the population making one choice can negatively affect the payoff for other choices. To extend our earlier example, this would allow for the expected percentage of religious believers in a population to reduce the payoffs associated

[6] The existence of self-consistent equilibria under these more general forms of endogenous social interactions is a consequence of Brouwer's fixed theorem in the same way as was the case for the initial multinomial logit model.

[7] In the previous models we have analyzed, J and J_l are allowed to take on negative values, but no cross-choice effects are present.

with agnosticism or atheism, whereas no cross-choice effects exist between these two possibilities.

The conditions for identification for the model defined by (17) and (18) with expectations described by (16) is very similar to that of Theorem 2. Formally, we have the following corollary.

Corollary 1. Identification for generalized multinomial logit model with social interactions

Suppose that individual choice is described by eq. (17) with self-consistent beliefs defined by eq. (16). Denote $m_{g(i),l} = p_{g(i),l} - p_{g(i),0}$. If, in addition to the assumptions found in Theorem 2, the support of the set of vectors $m_{g(i)} = \left(m_{g(i),1}, \ldots m_{g(i),L-1}\right)$ does not lie in a proper linear subspace of R^{L-1}, then the true set of model parameters $\left(k_1, c_1, d_1, J_1, \ldots, k_{L-1}, c_{L-1}, d_{L-1}, J_{L-1}\right)$ is identified relative to any distinct alternative.

Intuitively, the additional condition in the Corollary adds sufficient variability in aggregate choice probabilities to allow for identification of the individual elements of J_l. This additional variability allows us to mimic the proof of Theorem 2 and apply it to Corollary 1 as shown in the Technical Appendix.

Finally, it is worth noting that the multinomial and binary choice models contain an interesting difference with respect to the presence of zero restrictions on the model parameters. Unlike the binary choice model, for the multinomial choice model there may be zero restrictions on particular elements of k_l, c_l, d_l, J_l that apply to one choice but not another. This means, for example, that a variable that is relevant for two of the choices may be known to be irrelevant for the others. However, since choices are determined by payoff maximization as in (1), the absence of a regressor in the payoff for a given possibility does not mean that it is irrelevant to whether that possibility is chosen. This reasoning suggests that there may be ways to employ regressors that are omitted from given choice-specific payoffs to identify those choice parameters. This may also prove to be a route for finding choice-specific instrumental variables as needed in various forms of the model.

IV. Multinomial choice under alternative error assumptions

The basic logic of the multinomial model is straightforward to generalize. This can be seen if one considers the preference structure

$$V_{i,l} = h_{i,l} + Jp_{i,l}^{e} + \beta^{-1}\varepsilon_{i,l} \tag{19}$$

This is the same preference structure we worked with earlier, except that β is now explicitly used to index the intensity of choice (in the McFadden sense) rather than as a parameter of the distribution of the random payoff term $\varepsilon_{i,l}$. We assume that these unobserved utility terms are independent and identically distributed with a common distribution function $F_{\varepsilon}(\cdot)$.

For this model, the probability that agent i makes choice l is

$$\mu\begin{pmatrix} \varepsilon_{i,0} - \varepsilon_{i,l} \le \beta\left(h_{i,l} - h_{i,0}\right) + \beta J\left(p_{i,l}^{e} - p_{i,0}^{e}\right), \dots, \\ \varepsilon_{i,L-1} - \varepsilon_{i,l} \le \beta\left(h_{i,l} - h_{i,L-1}\right) + \beta J\left(p_{i,l}^{e} - p_{i,L-1}^{e}\right) \end{pmatrix} \tag{20}$$

Following Anderson, dePalma, and Thisse (1992, pg. 36), conditional on a realization of $\varepsilon_{i,l}$, the probability that l is chosen is

$$\prod_{j \ne i} F_{\varepsilon}\left(\beta h_{i,l} - \beta h_{i,j} + \beta J p_{i,l}^{e} - \beta J p_{i,j}^{e} + \varepsilon_{i,l}\right) \tag{21}$$

which immediately implies that the probability of the choice l without conditioning on the realization of $\varepsilon_{i,l}$ is

$$p_{i,l} = \int \prod_{j \ne l} F_{\varepsilon}\left(\beta h_{i,l} - \beta h_{i,j} + \beta J p_{i,l}^{e} - \beta J p_{i,j}^{e} + \varepsilon\right) dF_{\varepsilon}\ . \tag{22}$$

Eq. (22) provides a multinomial choice model whose structure is fully analogous to the multinomial logit structure developed in Sections II and III. Under self-consistency, the aggregate choice probabilities of this general multinomial choice model are the solutions to

$$p_l = \int\int\prod_{j\neq l} F_{\varepsilon}\left(\beta h_l - \beta h_j + \beta J p_l - \beta J p_j + \varepsilon\right) dF_{\varepsilon} dF_h \tag{23}$$

As in the multinomial logit case, the compound parameter βJ plays a critical role in determining the number of self-consistent equilibrium choice probabilities p_l. This finding is formalized in Theorem 3.

Theorem 3. Uniqueness versus multiplicity of self-consistent equilibria in multinomial choice models

Suppose that individual choices and associated self-consistent equilibria are described by (22) and (23). Assume that $h_{i,l} = 0 \;\; \forall i,l$ and $\varepsilon_{i,l}$ are independent across i and l. There exists a threshold T such that if $\beta J < T$, then there is a unique self-consistent equilibrium, whereas if $\beta J > T$ there exist at least three self-consistent equilibria.

The relationship between βJ and the number of equilibria is less precise than was found in Theorem 1, the multinomial logit case, as Theorem 3 does not specify anything about the way in which L, the number of available choices, affects the number of equilibria. This lack of precision is to be expected since we did not specify the distribution of the errors.

One can also develop an analog to the identification results we have obtained for the multinomial logit model. We will work with the same normalizations as used in the multinomial logit case and will again assume that $\varepsilon_{i,l}$ is independent of X_i and $Y_{g(i)}$ $\forall i,l$. Under self consistency, eq. (22) defines a continuous mapping (23) from the simplex

$$S =_{def} \left\{ \left(p_0, ..., p_{L-1} \right) \mid p_l \geq 0, j = 0, ..., L-1, \sum_l p_l = 1 \right\} \quad (24)$$

into itself. Assume that this mapping is globally one-to-one. This is a "high level" assumption in the sense that it is an assumption that is imposed on the choice probabilities; ideally it is preferable to place assumptions on the payoff function and show that such a condition holds. However, for our purposes, the assumption should not be regarded as too extreme as it holds for standard cases such as the multinomial logit.

Global invertibility provides a route to identification. Recall that nonidentification means that there exist two sets of parameters that produce the same choice probabilities $p_{i,l}$ and hence the same choice probability differences $p_{i,l} - p_{i,0}$. It is immediate that this invariance requires that if there exist two distinct sets of parameters $\left(k_1, c_1, d_1, J_1, \ldots, k_{L-1}, c_{L-1}, d_{L-1}, J_{L-1}\right)$ and $\left(\overline{k}_1, \overline{c}_1, \overline{d}_1, \overline{J}_1, \ldots, \overline{k}_{L-1}, \overline{c}_{L-1}, \overline{d}_{L-1}, \overline{J}_{L-1}\right)$ that are observationally equivalent in the sense that the individuals choice probabilities they induce are equal, that

$$k_l + c_l' X_{i,l} + d_l' Y_{g(i),l} + J\left(p_l - p_0\right) = \overline{k}_l + \overline{c}_l' X_{i,l} + \overline{d}_l' Y_{g(i),l} + \overline{J}\left(p_l - p_0\right). \quad (25)$$

Eq. (25) is the same condition that was analyzed in the proof of Theorem 2 (compare with (A.7) in the Technical Appendix). The proof of Theorem 2 can therefore be adapted step by step to this case, allowing us to state Theorem 4.

Theorem 4. General parametric identification for the multinomial choice model

Let the true data generating process be given by (17)-(21) with the normalization $k_0 = 0,\ c_0 = 0,\ d_0 = 0,\ J_0 = 0$ and $\beta=1$. Assume that the error distribution F_ε is known. Assume that the mapping defined by (23) is globally one-to-one. Then the true set of model parameters $\left(k_1, c_1, d_1, J_1, \ldots, k_{L-1}, c_{L-1}, d_{L-1}, J_{L-1}\right)$ is identified relative to any distinct alternative under the same assumptions *i...vii* as found in Theorem 2.

Taken together, Theorems 3 and 4 show that our basic analysis of social interactions using the multinomial logit model are not driven by the specific random payoff distribution that is assumed but rather stem from the underlying logic of the model.[8]

V. Group choice and behavior choice

Our analysis so far has treated groups as predetermined. For contexts such as ethnicity or gender this is presumably appropriate. However, in other contexts, such as residential neighborhoods, group memberships are themselves presumably influenced by the presence of social interactions effects. Hence a complete model of the role of social interactions on individual and group outcomes requires a joint description of both the process by which groups are formed and the subsequent behaviors they induce. As yet, the literature on social interactions has not fully developed this joint approach. In particular, analyses such as Glaeser, Sacerdote, and Scheinkman (1996) and Brock and Durlauf (2001a,b) that focus on the micro structure of social interactions using interacting particle systems methods, have treated the interaction structures under study as exogenous. In contrast, models such as Bénabou (1993,1996) and Durlauf (1996a,b) that have focused on the determinants of groups (in both cases neighborhoods) have been less concerned with the modeling of the structure of social interactions.

Further, the failure to account for the way groups form may have important econometric implications. As discussed in Brock and Durlauf (2001b), and Manski (2000) and Moffitt (2001), endogenous neighborhood choice has important implications for econometric implementation of models of neighborhood effects. Yet endogeneity of neighborhood memberships need not be an impediment to identifying neighborhood effects. Brock and Durlauf (2001b) in fact show, that self-selection into neighborhoods, when correctly specified, can facilitate identification via the creation of additional determinants of individual behavior in linear models and/or by inducing nonlinearities in individual behavior, each of which eliminates possible collinearity between contextual effects and endogenous effects.

---

[8] One limitation of Theorem 3 is that it assumes that the distribution function F_ε is known. We are currently exploring identification in the case where F_ε is unknown.

In this section, we outline two approaches for the integration of group determination and individual choice in the presence of social interactions. First, we consider the integration of group choices into a linear model of behavior. Second, we integrate group and behavioral choices into a common multinomial choice framework.[9] We will not derive these models from an explicit formulation of preferences as our goal is to characterize the probability structure of behavioral choices in the presence of endogenous group memberships.

i. linear in means models and endogenous group membership

One approach to integrating group choice and behavioral decisions may be developed by integrating group choice into a model in which the behavior obeys a linear model. Such models are quite common in the empirical literature on social interactions and have been studied by Brock and Durlauf (2001a,b), Manski (1993), and Moffitt (2001). Following the formulation in Brock and Durlauf (2001a,b), behavioral choices ω_i are continuous and are described by

$$\omega_i = k + c'X_i + d'Y_{g(i)} + Jm_{g(i)} + \varepsilon_i \tag{26}$$

Relative to the multinomial choice model of behaviors, a key difference in this specification is that the possible ω_i values are ordered. Suppose that each individual assigns to each group an overall "quality" measure

$$I^*_{i,g} = \gamma' Z_{i,g} + \nu_{i,g} \tag{27}$$

---

[9] Other approaches also appear promising in terms of understanding the interplay between social interactions and group formation for particular environments. For example, Ekelund, Heckman, and Nesheim (2001) show how prices associated with residential neighborhood memberships contain important information that may be used to uncover social interaction effects. Another important approach is due to Epple and Sieg (1999) who show how to develop implications for the distribution of families across communities in Tiebout-type environments.

where $Z_{i,g}$ denotes those observable characteristics of i that influence his evaluation of group g and $\nu_{i,g}$ denotes an unobservable individual-specific quality term. Individual i is assumed to be a member of the group with the highest $I^*_{i,g}$. We assume that $E\left(\varepsilon_i \middle| X_i, Y_g, Z_{i,g}\right) = 0$ and $E\left(\nu_{i,g} \middle| X_i, Y_g, Z_{i,g}\right) = 0$. $\forall i, g$. Also, we assume that the variance of ε_i, σ^2_ε, and the correlation between ε_i and $\nu_{i,g}$, ρ, are independent of group membership. This is more restrictive than the assumptions made in Lee (1983); we make this stronger assumption in order to avoid unnecessary complications.

The formulation we have described raises interesting econometric issues. Specifically, the model embodies two major issues that have been studied in the econometrics literature. First, eq. (26), known as the linear-in-means-model, has been shown to suffer from serious identification problems in the absence of endogenous group membership. Specifically, Manski (1993) has shown that if there is a one-to-one correspondence between X_i and $Y_{g(i)}$ among the independent variables that appear in (26), (i.e. $Y_{g(i)}$ is the average value of X_i within group g), the parameters in (26) are not identified. The reason for this is that under the Manski assumption, $m_{g(i)}$ is linearly dependent on $Y_{g(i)}$. Second, linear models with self-selection into groups have received a great deal of attention in the econometric literature because of the inconsistency of ordinary least squares estimates of (26). The basic problem with self-selection is that in such cases one needs to account for the possibility that $E(\varepsilon_i \mid i \in g) \neq 0$, a property that will hold if ε_i and $\nu_{i,g}$ are correlated.

Our goal in the subsequent discussion is to show how one can identify the parameters of the model we have described. The identification problem will be shown to revolve around the explicit incorporation of a self-selection correction into the behavioral equation (26). Heckman (1979) represents the seminal work in how to address the effects of this type of sample selection. Lee (1983) has developed an approach to dealing with self-selection that we employ here. We emphasize that our purpose is illustrative in that we demonstrate identifiability only under a particular set of parametric assumptions. However, the logic of our argument is more general than the case we study and can be adapted to alternative sets of assumptions. Also, it is important to note that Ioannides and Zabel (2002b) recognized previously that an argument in Brock and Durlauf (2001b) on the use of self-selection correction to achieve identification in models with two groups could be extended to multiple groups when group membership follows a multinomial logit framework. Our derivation differs from theirs in two respects. First, we employ an approach to selection correction developed by Lee (1983) rather than that due to Dubin and McFadden (1984); the relative merits of the two are discussed in Schmertmann (1994) and Vella (1997). Second, we analyze how the nonlinearity of a selection correction affects identification.[10]

We require two assumptions. First we assume $v_{i,g}$ is double exponentially distributed as in eq. (4). Then, following Lee (1983, pg. 511 eq. (3.6)) the distribution function $\Lambda_g(\cdot)$ is defined as

$$\Lambda_g(\upsilon) = \frac{\exp(\upsilon)}{\exp(\upsilon) + \sum_{j \neq g} \exp\left(\gamma' Z_{i,j}\right)} \tag{28}$$

where relative to (4) parameter β is normalized to equal 1. This is the function that appears in (30) below. This assumption therefore means that the group choices obey the multinomial logit model we have already developed. Second, we assume that $\varepsilon_{i,g}$ is normally distributed; we denote the density and distributions of the standardized normal, $N(0,1)$ as $\phi(\cdot)$ and $\Phi(\cdot)$ respectively.

These assumptions allow one to transform (26) in such a way as to produce a model that accounts for $E(\varepsilon_i \mid i \in g) \neq 0$. Following Lee (1983, pg. 511, eq. (3.7)), whose analysis extends the argument that underlies Heckman (1979), one may rewrite (26) as

---

[10]One may also consider issues raised by unobservables which do not involve self selection for the linear-in-means model. For example, Graham and Hahn (2003) study a version of (26) where k is replaced by k_g. They explore alternative GMM and instrumental variables methods to identify the parameters of (26). Brock and Durlauf (2001b) discuss routes to identification that, for example, use differencing within groups to eliminate k_g for this context.

$$\omega_i = k + c'X_i + d'Y_{g(i)} + Jm_{g(i)} - \rho\sigma_\varepsilon\varphi_{g(i)}(\gamma' Z_{i,g(i)}) + \xi_{i,g(i)} \tag{29}$$

where

$$\varphi_{g(i)}(\upsilon) = \phi\left(\frac{\Phi^{-1}(\Lambda_{g(i)}(\upsilon))}{\Lambda_{g(i)}(\upsilon)}\right) \tag{30}$$

The function $\varphi_{g(i)}$ is ungainly, but is invaluable in terms of identification. In fact, there are two routes to identification in the model that are facilitated by the selection correction. To see this, it is easiest to follow Manski's assumption on the relationship between X_i and $Y_{g(i)}$ and consider

$$\omega_i = k + c'X_{i,g(i)} + d'X_{g(i)} + Jm_{g(i)} - \rho\sigma_\varepsilon\varphi(\gamma' Z_{i,g(i)}) + \xi_{i,g(i)} \tag{31}$$

If $\rho = 0$, then this model is not identified. In contrast, suppose that $\rho \neq 0$ and that $m_{g(i)}$ is not an element of $Z_{i,g(i)}$. In this case $\varphi_{g(i)}(\gamma' Z_{i,g(i)})$ is an individual-specific variable whose group level average does not appear in (30). As shown in Brock and Durlauf (2001a, Theorem 6), the presence of such a regressor means that identification of the regression parameters in (30) is possible.[11] Alternatively, suppose that $Z_{i,g(i)} = m_{i,g(i)}$, so that (outside unobserved heterogeneity), the only variable that influences group choices is the expected average behavior within the neighborhood. In this case, (30) is now a nonlinear in means model, in the sense that ω_i is linearly related to $Jm_{g(i)} - \rho\sigma_\varepsilon\varphi_{g(i)}(\gamma m_{g(i)})$. Brock and Durlauf (2001b, Theorem 7) show that nonlinear in means models of this type are locally identified, except for "hairline" cases.

[11] The condition is necessary, rather than sufficient, but the presence of the variable breaks the necessary linear dependence of $m_{g(i)}$ on $Y_{g(i)}$

Intuitively, the nonlinear relationship between ω_i and $m_{g(i)}$ precludes $m_{g(i)}$ from being linearly dependent when $Y_{g(i)} = X_{g(i)}$.[12]

This argument thus generalizes the analysis of identification and self-selection found in Brock and Durlauf (2001b, pp. 3328-3331). The key message for empirical work is that self-selection, if properly accounted for, can facilitate the identification of social interactions.

ii. a nested choice approach to integration behaviors and group memberships

A second approach to endogenizing group memberships may be developed using the nested logit framework originated by Ben Akiva (1973) and McFadden (1978). The basic idea of this framework is the following. An individual is assumed to make a joint decision of a group $g \in \{0,...G-1\}$ and a behavior $l \in \{0,...L-1\}$. We will denote the group choice of i as δ_i. The structure of this joint decision is nested in the sense that the choices are assumed to have a structure that allows one to decompose the decisions as occurring in two stages: first, the group is chosen and then the behavior.

The key feature of this type of model is the assumption that choices at each stage obey a multinomial logit probability structure. For the behavioral choice, this means that

$$\mu\left(\omega_i = l \middle| h_{i,l,g}, p^e_{i,l,g}, \delta_i = g\right) = \frac{\exp \beta\left(h_{i,l,g} + Jp^e_{i,l,g}\right)}{\sum_{j=0}^{L-1} \exp \beta\left(h_{i,l,g} + Jp^e_{i,l,g}\right)} \tag{32}$$

which is the same behavioral specification as (6). Group choices are somewhat more complicated. In the nested logit model, group choices are assumed to obey

---

[12]While this nonlinearity argument holds in principle, a common concern in empirical work with selection corrections is the "quality" of the identification for the range of observed data when identification is based on a nonlinearity argument, cf. Vella (1998, pg. 135). Hence, for the model we have described, the presence of an additional z_i that is not linearly dependent on X_i or $Y_{g(i)}$ may be very helpful in practice.

$$\mu\left(i \in g \middle| h_{i,l,g}, p^e_{i,l,g} \forall l, g\right) = \frac{\exp(\beta_g Z_{i,g})}{\sum_g \exp(\beta_g Z_{i,g})} \tag{33}$$

where

$$Z_{i,g} = E(\max_l h_{i,l,g} + Jp^e_{i,l,g} + \varepsilon_{i,l,g}) \tag{34}$$

with $\varepsilon_{i,l,g}$ independent and doubly exponentially distributed random variables across i and l for a given g. A standard result (e.g. Anderson, de Palma and Thisse (1992, pg. 46)) is that

$$E\left(\max\left(h_{i,l,g} + Jp^e_{i,l} + \varepsilon_{i,l,g} \middle| h_{i,l,g}, p^e_{i,l,g} \forall l, g\right)\right) = \beta^{-1} \log\left(\sum_l \exp \beta\left(h_{i,l,g} + Jp^e_{i,l,g}\right)\right) \tag{35}$$

Combining, (31)-(34), the joint group membership and behavior probabilities for an individual are thus described by

$$\mu\left(\omega_i = l, \delta_i = g \middle| h_{i,l,g}, p^e_{i,l,g} \forall l, g\right) =$$
$$\frac{\exp\left(\beta_g \beta^{-1} \log\left(\sum_l \exp \beta\left(h_{i,l,g} + Jp^e_{i,l,g}\right)\right)\right)}{\sum_g \exp\left(\beta_g \beta^{-1} \log\left(\sum_l \exp \beta\left(h_{i,l,g} + Jp^e_{i,l,g}\right)\right)\right)} \cdot \frac{\exp \beta\left(h_{i,l,g} + Jp^e_{i,l,g}\right)}{\sum_{j=0}^{L-1} \exp \beta\left(h_{i,l,g} + Jp^e_{i,l,g}\right)} \tag{36}$$

This probabilistic description may be faulted in that (35) is not directly derived from a utility maximization problem. In fact, a number of papers have identified conditions under which (35) is consistent with utility maximization, cf. McFadden (1978) and Borsch-Supan (1990) for discussion. A simple condition (cf. Anderson, dePalma, and Thisse, 1992, pg. 48) that renders (35) compatible with a well posed utility maximization problem is $\beta_g \leq \beta$, which in essence requires that the dispersion of random payoff terms across groups is lower than the dispersion in random payoff terms across behavioral choices within a group.

There has yet to be any analysis of models such as (35) when self-consistency is imposed on the expected group choice percentages $p^e_{i,l,g}$. Such an analysis should provide a number of interesting results. For example, a nested structure of this type introduces a new mechanism by which multiple equilibria may emerge, namely the influence of beliefs about group behaviors on group memberships, which reciprocally will affect behaviors. This additional channel for social interactions, in turn, raises new identification questions.

VI. Conclusions

This paper has described an approach to modeling social interactions that extends standard tools in the discrete choice literature, namely logit models of choice. The approach allows for the incorporation of a range of alternative types of social interactions into individual decisionmaking in a way that retains the logic of economic behavior while at the same time provides additional richness to the determinants of individual behavior. A virtue of the approach is that the theoretical model can be directly taken to data, both in the sense that the description of equilibrium choices is simultaneously a likelihood function and because the various group influences embedded in the model are identifiable under intuitive and reasonably weak conditions. This has been demonstrated through the analysis of a leading case, namely, a multinomial logit version of the model. We have also shown that the qualitative theoretical and econometric features of our leading case, the multinomial logit model, also apply to alternative formulations of the random payoff process. Finally, we have illustrated how one can integrate choices about group memberships with choices on behaviors using a nested multinomial logit model.

More generally, we believe that there is wide scope for the better integration of sociological ideas and economic reasoning to provide a deeper understanding of the various phenomena that engage both disciplines. An important feature of the new social economics (Durlauf and Young (2001)) is that it attempts to take account of phenomena ranging from crime to fertility to education where sociological factors would seem to play a key role. Economists have long understood the importance of addressing such factors. Arrow (1974), for example, remarks

"…there are profound difficulties with the price system, even, so to speak, within its own logic, and these strengthen the view that, valuable though it is in certain realms, it cannot be made the arbiter of social life," (pg. 21-22)

The models we analyze address one aspect of the general issues raised by Arrow and others by embedding individual choice in contexts where social factors exist outside the realm of markets or prices. In turn, we believe that the choice-based approach we have developed is valuable in terms of providing a logical structure to sociological-style arguments. One reason for this judgment is that social explanations of aggregate phenomena are most useful when the implied rules for individual behavior are interpretable as purposeful decisions. Arrow (1994) makes precisely this argument:

"It is a salutary check on any theory of the economy or any other part of society that the explanations make sense on the basis of the individuals involved." (pg.3)

The theoretical and econometric approach we advocate is inspired by and attempts to implement Arrow's vision.

Technical Appendix

Proof of Theorem 1

In verifying this theorem, it is convenient to rewrite eq. (8) so as to measure the deviation of choice probabilities from $l=0$; i.e. we work with $m_l = p_l - p_0$ and $g_{i,l} = h_{i,l} - h_{i,0}$, $l = 1 \ldots L-1$. The probability differences m_l may be written as

$$m_l = \int \left(\exp\left(\beta g_{i,l} + \beta J m_l \right) - 1 \right) / W_i dF_h \tag{A.1}$$

where

$$W_i = \sum_{l=1}^{L-1} \exp\left(\beta g_{i,l} + \beta J m_l \right) + 1 \tag{A.2}$$

Letting $m = (m_1, \ldots, m_{L-1})$ and $g = (g_1, \ldots, g_{L-1})$, the $L-1$ equations defined by (A.1) and (A.2) constitute a mapping from $[-1,1]^{L-1}$ to $[-1,1]^{L-1}$ which we denote as $\psi(m, \beta J, g)$. Fixed points of the mapping are defined by $m = \psi(m, \beta J, g)$ and constitute the self-consistent equilibria of the model. The question of the relationship between the behavioral parameters of the model and the number of equilibria may be answered by determining how this mapping changes as βJ changes under the assumptions of the theorem.

Under the assumption that $h_{i,l} = k \ \ \forall i, l$, it is of course the case that $g = 0$, since there are no differences in the private deterministic utility differences between choices. This assumption allows the analysis to focus entirely on the effect of βJ. Under this assumption, there exists a fixed point $m = 0$ for any value of βJ. To see whether other fixed points exist, we compute the derivative of $\psi(m, \beta J, 0)$ with respect to m at the fixed point $m = 0$. The Jacobian matrix of derivatives of $\psi(m, \beta J, 0)$ taken with respect to elements of m contains diagonal elements

$$\frac{\partial \psi_l\left(m,\beta J,0\right)}{\partial m_l}=\frac{\beta J\exp\left(\beta J m_l\right)}{\left(\sum_{i=1}^{L-1}\exp\left(\beta J m_j\right)\right)+1}-\frac{\beta J\left(\exp\left(\beta J m_l\right)-1\right)\left(\exp\left(\beta J m_l\right)\right)}{\left(\left(\sum_{i=1}^{L-1}\exp\left(\beta J m_j\right)\right)+1\right)^2} \tag{A.3}$$

and off-diagonal elements

$$\frac{\partial \psi_l\left(m,\beta J,0\right)}{\partial m_j}=-\frac{\beta J\left(\exp\left(\beta J m_l\right)-1\right)\left(\exp\left(\beta J m_j\right)\right)}{\left(\left(\sum_{i=1}^{L-1}\exp\left(\beta J m_i\right)\right)+1\right)^2} \tag{A.4}$$

so for $m=0$,

$$\frac{\partial \psi_l\left(0,\beta J,0\right)}{\partial m_k}=\frac{\beta J}{L}\ \textit{if}\ l=k, 0\ \textit{otherwise} \tag{A.5}$$

Consider the set of vectors m of the form $\left(m_1,0\right)$, i.e. vectors with zero components except for the first element. Denote the set of all such vectors by Γ_1. Observe that Γ_1, which lies in R^{L-1}, is an invariant set with respect to $\psi\left(m,\beta J,0\right)$ as each element of Γ_1 maps onto an element of Γ_1. We now focus on the first component of the ψ map, ψ_1, which can always be written as a one-dimensional map on R, denote this as $\rho\left(m_1,\beta J\right)$.

Finally, consider fixed points for the mapping $\rho\left(m_1,\beta J\right)$. One fixed point exists, as previously observed, at $m_1=0$. Further, recall that by (A.5),

$$\frac{\partial \rho\left(0,\beta J\right)}{\partial m_1}=\frac{\beta J}{L} \tag{A.6}$$

It is easy to verify that ρ is a convexo-concave function with respect to its first argument. This means that as $\frac{\beta J}{L}$ becomes greater than 1, two new fixed points must emerge. This argument is sufficient to verify Theorem 1.

Proof of Theorem 2

This proof is a generalization of the proof for identification for a binary choice model in Brock and Durlauf (2001a,b) which in turns develops a mode of argument found in Manski (1988). Suppose that for $\left(k_1, c_1, d_1, J_1, \ldots, k_{L-1}, c_{L-1}, d_{L-1}, J_{L-1}\right)$, the set of true parameters for the multinomial choice model, there exists another vector $\left(\bar{k}_1, \bar{c}_1, \bar{d}_1, \bar{J}_1, \ldots, \bar{k}_{L-1}, \bar{c}_{L-1}, \bar{d}_{L-1}, \bar{J}_{L-1}\right)$ that generates the same observed data. If both sets of parameters generate the same probabilities for the observables, this implies,

$$\ln\left(p_{li} / p_{0i}\right) = k_l + c_l' X_{i,l} + d_l' Y_{g(i),l} + J_l\left(p_l - p_0\right) = \bar{k}_l + \bar{c}_l' X_{i,l} + \bar{d}_l' Y_{g(i),l} + \bar{J}_l\left(p_l - p_0\right) \qquad \text{(A.7)}$$

for $l = 1, \ldots, L-1$. From assumption *iv* of the Theorem, there is at least one neighborhood for each choice l such that within that neighborhood, $X_{i,l}$ is not contained in a proper linear subspace of R^r. Hence, (A.7) can hold if and only if $c_l = \bar{c}_l$. This argument applies to each of the possible choices, so $c_1, \ldots, c_{L-1}$ are identified.

Given that $c_1, \ldots, c_{L-1}$ are identified, it must be the case that the Theorem is true if $J = \bar{J}$; lack of identification would imply that either 1) $\left(X_{i,l}, Y_{g(i),l}\right)$ lies in a proper linear subspace of R^{r+s}, which would violate assumption *i* of the theorem or 2) that some linear combination of elements in $\left(X_{i,l}, Y_{g(i),l}\right)$ is constant, which would violate assumption *iii*. We can therefore restrict attention to the case $J \neq \bar{J}$. Define $m_{g(i),l} = p_{g(i),l} - p_{g(i),0}$. Notice that $m_{g(i),l}$ is bounded between -1 and 1. Since $c_1, \ldots, c_{L-1}$ are identified, (A.7) requires that

$$k_l - \bar{k}_l + \left(d_l' - \bar{d}_l'\right) Y_{g(i),l} = \left(\bar{J}_l - J_l\right) m_{g(i),l} \tag{A.8}$$

since $c_l = \bar{c}_l \; \forall l$. Since $m_{l,g(i)}$ cannot be zero for all $g(i)$ by assumption *vi* and since assumption *v* implies that $\left(d_l' - \bar{d}_l'\right) Y_{g(i),l}$ is unbounded if $d_l \neq \bar{d}_l$, we have a contradiction to the boundedness of $\left(\bar{J}_l - J_l\right) m_{g(i),l}$ unless $d_l = \bar{d}_l$.

Failure of identification now requires that

$$k_l - \bar{k}_l = \left(\bar{J}_l - J_l\right) m_{g(i),l} \tag{A.9}$$

holds across all groups. Given assumption *vi*, the nonconstancy of $p_{l,n(i)}$ (and thus $m_{g(i),l}$) across groups, can only hold across neighborhoods if $J_l = \bar{J}_l$. Substituting this into (A.7), it is obvious that $k_l = \bar{k}_l$ which completes the proof.

Proof of Corollary 1.

Following the proof of Theorem 2, Corollary 1 will be proved if one can show that

$$\ln\left(p_{li} / p_{0i}\right) = k_l + c_l' X_{i,l} + d_l' Y_{g(i),l} + J_l'\left(p_l - p_0\right) = \bar{k}_l + \bar{c}_l' X_{i,l} + \bar{d}_l' Y_{g(i),l} + \bar{J}_l'\left(p_l - p_0\right) \tag{A.10}$$

cannot hold for any $\left(\bar{k}_1, \bar{c}_1, \bar{d}_1, \bar{J}_1, \ldots, \bar{k}_{L-1}, \bar{c}_{L-1}, \bar{d}_{L-1}, \bar{J}_{L-1}\right)$ distinct from $\left(k_1, c_1, d_1, J_1, \ldots, k_{L-1}, c_{L-1}, d_{L-1}, J_{L-1}\right)$. The argument made immediately after (A.7) applies to (A.10) as well, which means that $c_1, \ldots, c_{L-1}$ are identified**.** Similarly, the argument made after (A.8) applies to (A.10) which implies that $d_1, \ldots, d_{L-1}$ are identified. Hence, we can restrict our attention to

$$k_l - \bar{k}_l = \left(\bar{J}_l - J_l\right)' m_{g(i)} \tag{A.11}$$

Given the assumption of the Corollary that the support of $m_{g(i)}$ does not lie in a proper linear subspace of R^{L-1} (A.11) can only hold if $\bar{J}_l = J_l$. This implies $k_l = \bar{k}_l$ which verifies the Corollary.

Proof of Theorem 3

To verify Theorem 3, we follow the same logic as the proof for Theorem 1. Define the mapping

$$m_l = \int \left(p_{i,l} - p_{i,0} \right) dF_h \tag{A.12}$$

$l = 1 \ldots L-1$. By eq. (23) in the text, this defines a mapping $\psi\left(m_1, 0, \beta J, \beta\right)$ from m to m. Under the assumption that $h_{i,l} = 0 \quad \forall i,l$, it is straightforward to verify that $\Gamma_1 = \left(m_1, 0\right)$ is an invariant set under this mapping. Hence, in parallel to the proof of Theorem 1, there is a mapping $\rho\left(m_1, \beta J, \beta\right)$ from R to R such that

$$\rho\left(m_1, \beta J, \beta\right) = \psi_1\left(m_1, 0, \beta J, \beta\right) \tag{A.13}$$

Under the assumptions that $h_{i,l} = 0 \quad \forall i,l$ and $\varepsilon_{i,l}$ are independent across i and l, it is immediate that $m_1 = 0$ must be a fixed point of this mapping. The existence of other fixed points will depend on the derivative of $\rho\left(\cdot,\cdot,\cdot\right)$ at $m_1 = 0$.

To analyze this derivative, note that

$$
\begin{aligned}
p_{i,1} - p_{i,0} &= \int \prod_{j \neq 1} F_\varepsilon \left(\beta J m_1 - \beta J m_j + \varepsilon \right) dF_\varepsilon - \int \prod_{j \neq 0} F_\varepsilon \left(\beta J m_0 - \beta J m_j + \varepsilon \right) dF_\varepsilon = \\
&\int \prod_{j \neq 1} F_\varepsilon \left(\beta J m_1 + \varepsilon \right) dF_\varepsilon - \int \prod_{j \neq 0} F_\varepsilon \left(-\beta J m_1 + \varepsilon \right) dF_\varepsilon
\end{aligned}
\tag{A.14}
$$

since $m_l = 0$ for $l = 2, \ldots, L-1$. Further, given the assumptions that $h_{i,l} = 0 \;\; \forall i, l$ and $\varepsilon_{i,l}$ are independent across i and l, $p_{i,1} - p_{i,0} = m_1$. Therefore, we can define a map A from m_1 to itself such that

$$
m_1 = A\left(\beta J m_1, L, \beta \right) \tag{A.15}
$$

This function, which is clearly monotonic and bounded between -1 and 1, depends on L through the products in (A.12) via (A.14). The derivative of this function is $\beta J A'\left(\beta J m_1, L, \beta \right)$. Consider $\beta J A'\left(0, L, \beta \right)$, the derivative of the function at the fixed point $m_1 = 0$. Following the same argument in the proof of Theorem 1, if $\beta J A'\left(0, L, \beta \right) < 1$, then the fixed point $m_1 = 0$ is unique whereas if $\beta J A'\left(0, L, \beta \right) > 1$ then at least two additional fixed points must exist. Hence the magnitude of βJ can be varied so as to move from a unique to multiple (at least three) equilibria. This verifies Theorem 3.

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.*

Amaro de Matos, J. and J. Perez, (1991), "Fluctuations in the Curie-Weiss Version of the Random Field Ising Model," *Journal of Statistical Physics*, 62, 587-608.

Anderson, S., A. de Palma, and J. Thisse, (1992), *Discrete Choice theory of Product Differentiation*, Cambridge: MIT Press.

Arrow, K. (1974), *The Limits of Organization*, New York: W. W. Norton.

Arrow, K., (1994), "Methodological Individualism and Social Knowledge," *American Economic Review*, 84, 2, 1-9.

Arrow, K. and F. Hahn, (1971), *General Competitive Analysis*, Amsterdam: North Holland.

Bayer, P. and Timmins, C., (2001), "Identifying Social Interactions in Endogenous Sorting Models," mimeo, Yale University.

Becker, G. and K. Murphy, (2000), *Social Economics*, Cambridge: Harvard University Press.

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

Bénabou, R., (1996), "Equity and Efficiency in Human Capital Investment: The Local Connection," *Review of Economic Studies*, 62, 237-264.

Ben Akiva, M., (1973), *Structure of Passenger Travel Demand Models*, Ph.D dissertation, Department of Civil Engineering, MIT.

Ben Akiva, M. and S. Lerman, (1985), *Discrete Choice Analysis: Theory and Application to Travel Demand*, Cambridge: MIT Press.

Borsch-Supan, A., (1990), "On the Compatibility of Nested Logit Models with Utility Maximization," *Journal of Econometrics*, 43, 373-388.

Brewster, K. (1994), "Race Differences in Sexual Activity Among Adolescent Women: The Role of Neighborhood Characteristics," *American Sociological Review*, 59, 408-424.

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

Brock, W. and Durlauf, S., (1999), “A Formal Model of Theory Choice in Science,” *Economic Theory*, 14, 113-130.

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

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

Brock, W. and Durlauf, S., (2002), “A Multinomial Choice Model of Neighborhood Effects,“ *American Economic Review*, 92, 298-303.

Brooks-Gunn, J., G. Duncan, P. Klebanov, and N. Sealand, “Do Neighborhoods Affect Child and Adolescent Development?,” *American Journal of Sociology*, 00, 353-395.

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

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

Corcoran, M., R. Gordon, D. Laren, and G. Solon, (1992), “The Association between Men's Economic Status and Their Family and Community Origins,” J*ournal of Human Resources*, 27, 575-601.

Costa, D. and M. Kahn, (2001), “Cowards and Heroes: Group Loyalty in the Civil War,” *National Bureau of Economic Research Working Paper 8627*.

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

Datcher, L., (1982), “Effects of Community and Family Background on Achievement,” *Review of Economics and Statistics*, 64, 32-41.

Dubin, J. and D. McFadden, (1984), “An Econometric Analysis of Residential Electric Appliance Holdings and Consumption,” *Econometrica*, 52, 345-362.

Durlauf S., (1996a), “A Theory of Persistent Income Inequality,” *Journal of Economic Growth*, 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. and H. P. Young, (2001), “The New Social Economics,” in *Social Dynamics*, S. Durlauf and H. P. Young, eds., Cambridge: MIT Press, 1-14.

Ekelund, I., Heckman, J. and L. Nesheim, (2001), "Identification and Estimation of Hedonic Price Models," mimeo, Department of Econoimcs, University of Chicago.

Ellis, R., (1985), *Entropy, Large Deviations, and Statistical Mechanics*, New York: Springer-Verlag.

Ellis, R. and K. Wang, (1990), "Limit Theorems for the Empirical Vector of the Cuire-Weiss-Potts Model," *Stochastic Processes and Their Applications*, 40, 251-288.

Epple, D. and H. Sieg, (1999), "Estimation Equilibrium Models of Local Jurisdictions," *Journal of Political Economy*, 107, 645-681.

Ginther, D., R. Haveman, and B. Wolfe, (2000), "Neighborhood Attributes as Determinants of Children's Outcomes: How Robust are the Relationships?" *Journal of Human Resources*, 35, 4, 603-642.

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

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

Heckman, J. (1979), "Sample Selection Bias as a Specification Error," *Econometrica*, 47, 153-161.

Hoff, K., and A. Sen, (2000), "Home-Ownership, Community Interactions, and Segregation," mimeo, World Bank.

Ioannides, Y. and J. Zabel, (2002a), "Neighborhood Effects and Housing Demand," mimeo, Tufts University and forthcoming, *Journal of Applied Econometrics*.

Ioannides, Y. and J. Zabel, (2002b), "Interactions, Neighborhood Selection, and Housing Demand," mimeo, Department of Economics, Tufts University.

Katz, L., J. Kling and J. Liebman, (2001), "Moving to Opportunity in Boston: Early Results of a Randomized Mobility Experiment," *Quarterly Journal of Economics*, CXVI, 607-654.

Kooreman, P. and A. Soetevent, (2002), "A Discrete Choice Model with Social Interacitons: An Analysis of High School Teen Behavior," mimeo, Department of Economics, University of Groningen.

Lee, L.-F., (1983), "Generalized Econometric Models with Selectivity," *Econometrica*, 51, 507-512.

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

Lundberg, S. and R. Startz, (1998), "On the Persistence of Racial Inequality," *Journal of Labor Economics*, 16, 2, 292-323.

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

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

McFadden, D., (1978), "Modelling the Choice of Residential Location," in A. Karlvist, L. Lunqvuist, F. Snickars, and J. Weibull, eds., *Spatial Interaction Theory and Planning Models*, Amsterdam: North Holland, 75-96.

McFadden, D. (1984) "Qualitative Response Models," in Z. Griliches and M. Intriligator, eds., *Handbook of Econometrics 2*, Amsterdam: North Holland, 1395-1457.

Minkin, A., (2002), "Heterogeneous Social Interactions," mimeo, Department of Economics, University of Wisconsin.

Moffitt, R., (2001), "Policy Interventions, Low-Level Equilibria, and Social Interactions," in S. Durlauf and H. P. Young, eds., *Social Dynamics*, Cambridge: MIT Press, 45-82.

Nigmatullin, E., (2002), "Neighborhood Interactions and Fertility Dynamics," mimeo, Department of Economics, University of Wisconsin.

Plotnick, R. and S. Hoffman, (1999), "The Effect of Neighborhood Characteristics on Young Adult Outcomes: Alternative Estimates," *Social Science Quarterly*, 80, 1-18.

Rosenbaum, E. and L. Harris, (2001), "Residential Mobility and Opportunities: Early Impacts of the Moving to Opportunity Demonstration in Chicago," *Housing Policy Debate*, 12, 321-346.

Rosenbaum, J., (1995), "Changing the Geography of Opportunity by Expanding Residential Choice: Lessons from the Gautreaux Program," *Housing Policy Debate*, 6, 231-269.

Rosenbaum, J. and S. Popkin, (1991), "Employment and Earnings of Low-Income Blacks Who Move to Middle Class Suburbs," in *The Urban Underclass*, C. Jencks and P. Peterson, eds., Washington DC: Brookings Institution Press.

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

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

Schmertmann, C., (1994), "Selectivity Bias Correction Methods in Polychotomous Sample Selection Models," *Journal of Econometrics*, 60, 101-132.

Sirakaya, S., (2002), "Recidivism and Social Interactions," mimeo, Department of Economics, University of Wisconsin.

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

South, S. and K. Crowder, (1999), "Neighborhood Effects on Family Formation: Concentrated Poverty and Beyond," *American Sociological Review*, 64, 113-132.

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

Vella, F., (1998), "Estimating Models with Sample Selection Bias: A Survey," *Journal of Human Resources*, 33, 1, 127-169.