# Interactions-Based Models

William A. Brock
Steven N. Durlauf

SFI WORKING PAPER: 2000-05-028

SFI Working Papers contain accounts of scientific work of the author(s) and do not necessarily represent the views of the Santa Fe Institute. We accept papers intended for publication in peer-reviewed journals or proceedings volumes, but not papers that have already appeared in print. Except for papers by our external faculty, papers must be based on work done at SFI, inspired by an invited visit to or collaboration at SFI, or funded by an SFI grant.

©NOTICE: This working paper is included by permission of the contributing author(s) as a means to ensure timely distribution of the scholarly and technical work on a non-commercial basis. Copyright and all rights therein are maintained by the author(s). It is understood that all persons copying this information will adhere to the terms and constraints invoked by each author's copyright. These works may be reposted only with the explicit permission of the copyright holder.

www.santafe.edu

**SANTA FE INSTITUTE**

Interactions-Based Models


## Abstract

This paper describes a range of methods which have been proposed to study interactions in economic and social contexts. By interactions, we refer to interdependences between individual decisions which are not mediated by markets. These types of models have been employed to understand phenomena ranging from the effect of neighborhoods on the life prospects of children to the evolution of political party platforms. We provide a general choice-based framework for modelling such interactions which subsumes a number of specific models which have been studied. This framework illustrates the relationship between interactions-based models and models in statistical mechanics. Our analysis is then extended to the econometrics of these models, with an emphasis on the identification of group-level influences on individual behavior. Finally, we review some of the empirical work on interactions which has appeared in the social science literature.



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

# INTERACTIONS-BASED MODELS

William A. Brock

Department of Economics, University of Wisconsin

Steven N. Durlauf

Department of Economics, University of Wisconsin

Revised Draft: January 27, 2000

This paper has been prepared for the Handbook of Econometrics 5, James J. Heckman and Edward Leamer, eds. This work has benefited from our interactions over the years with Kenneth Arrow, Lawrence Blume, James Heckman, Charles Manski, Peter Phillips and H. Peyton Young. We thank seminar participants at the Brookings Institution, Michigan State University, NBER, Northwestern University, University of Geneva, University of Lausanne, University of Michigan, University of Texas and the University of Wisconsin as well as Kenneth Arrow, Arthur Goldberger, Andros Kourtellos, Artur Minkin, Derek Neal, Eldar Nigmatullin, and the MacArthur Research Network on Social Interactions and Economic Inequality for providing helpful comments on various drafts of this paper. Financial support from the John D. and Catherine T. MacArthur Foundation, National Science Foundation, Romnes Fund, and Vilas Trust is gratefully acknowledged.

## 1. Introduction

"The principal task of the social sciences lies in the explanation of social phenomena, not the behavior of single individuals. In isolated cases the social phenomena may derive directly, through summation, from the behavior of individuals, but more often this is not so. Consequently, the focus must be on the social system whose behavior is to be explained. This may be as small as a dyad or as large as a society or even a world system, but the essential requirement is that the explanatory focus be on the system as a unit, not on the individuals or other components which make it up." (pg. 2)

"(A)n internal analysis based on actions and orientations of units at a lower level can be regarded as more fundamental, constituting more nearly a theory of system behavior, than an explanation which remains at the system level...although an explanation which explains the behavior of a social system by the actions and orientations of some entities between the system level and the individual level may be adequate for the purpose at hand, a more fundamental explanation based upon the actions and orientations of individuals is generally more satisfactory." (pg. 4)

James Coleman (1990)

The role of interactions in economic outcomes has become an important area of research over the last decade. By interactions-based models, we refer to a class of economic environments in which the payoff function of a given agent takes as direct arguments the choices of other agents. The goal of such an analysis is to provide an explanation of group behavior which emerges from the interdependences across individuals.

In some respects, interactions-based models would appear to be nothing but a variant of game-theoretic formulations of decisionmaking; see Blume (1997) and Young (1998) for an excellent syntheses of a number of game-theoretic models

from the interactions perspective and Morris (1998) for a game-theoretic analysis of interaction structures. Further, Jones (1984), Cooper and John (1988) and Milgrom and Roberts (1990), there has been a great deal of work explicitly focusing on how one type of interaction effects, complementarities, can lead to multiple equilibria and other interesting aggregate phenomena, including breakdowns of the law of large numbers (Jovanovic (1985)). Indeed, following Bryant (1985), macroeconomic models of complementarities have become a standard research tool. Similarly, analyses such as Bernheim (1994) have shown how conformity effects can produce customs, fads and highly different subcultures within a given population.

Similarly, social sciences other than economics have a much longer tradition of looking for interaction effects. One particularly important example is the Coleman Report of 1966 (Coleman, *et al*, 1966), which argued that school performance of the disadvantaged was much more amenable to improvement through manipulation of peer group influence than by increased per student expenditures. While the Coleman Report itself has not withstood subsequent scrutiny, its impact on both social science research and public policy was and is immense (see Heckman and Neal (1996) for discussion.) See Blalock (1984) for additional discussion of sociological approaches. Another example is linguistics, where the role of interactions in influencing dialect choice has been well understood for decades (cf. Labov (1972a,b)).

What distinguishes the new research on interactions-based models is the explicit attention given to formulating how each individual's behavior is a function of the characteristics or behavior of others and then studying what aggregate properties emerge in the population. This approach typically, though not always, is done in the form of first specifying a conditional probability measure which describes each individual's behavior as a function of the rest of the population and then determining what joint probability measures are compatible with these conditional measures. This particular approach means that interactions-based approaches have typically been deeply reliant on the use of the

probability theory which underlies statistical mechanics methods in physics. (Mathematicians generally refer to statistical mechanics models as interacting particle systems. These models also fall into the broader class of probability models known as random fields.) The value of this approach is that it permits one to specify individual and social aspects of behavior simultaneously, and thereby address aggregate behavior in a way consistent with the sort of methodological individualism advocated by Coleman.

Interactions-based models have been applied to a wide range of contexts both within economics and within social science more generally. A sense of this range can be given through an admittedly incomplete survey of applications; see Durlauf (1997), Kirman (1997), and Rosser (1999) for additional overviews.

## 1. Neighborhoods and inequality

Much of the recent literature on persistent income inequality has focused on the role of neighborhood influences on socioeconomic outcomes. Theoretical models such as Bénabou (1993,1996a,b), Cooper (1998), Durlauf (1996a,b) share a common assumption that individual human capital acquisition depends on the behaviors and/or characteristics of community members. These influences may range from peer group effects, in which the costs to one person from investing effort in education are decreasing in the effort levels of others (Bénabou (1993)), to role model effects, in which the aspirations of a student are affected by the observed education/occupation outcomes among adults in his community (Streufert (1991)), to labor market connections (Granovetter (1995), Montgomery (1991,1992)), in which the probability with which one makes a successful job match depends on the information possessed by members of one's social network. Similar types of spillovers were used much earlier in Loury (1977) to provide a theory of racial income differences. Examples of studies which have adduced empirical evidence of neighborhood effects include Crane (1991), Case and Katz (1991), Haveman and Wolfe (1995). Within the psychology literature, there is

rich evidence on the importance of peer group effects, as illustrated, for example, in Brown (1990) and Brown, Clasen, and Eicher (1986). Finally, recent work by Casella and Rauch (1997,1998) shows how ethnic social networks can influence patterns of international trade through similar mechanisms with attendant implications for ethnic patterns of inequality.

## 2. Spatial agglomeration

The role of interactions effects in determining location decisions has been analyzed in many contexts. Schelling's (1971) work on racial segregation, illustrates how weak preferences by individuals for neighbors of similar ethnicity can lead to complete segregation. This work is possibly the first interactions-based model to be studied in the social sciences; see Granovetter and Soong (1988) for a number of extensions and generalizations of this original framework. Arthur (1987) has shown how sequential locational decisions, combined with locational spillover effects, can produce agglomerations of economic activity such as the Silicon Valley. Similar models, with a richer microeconomic structure, have been subsequently analyzed by Krugman (1996). In related work, Kelly (1995) has illustrated the evolution of geographically defined trade networks.

## 3. Technology choice

The adoption of particular technological standards is a well-studied case both by economic historians and economic theorists. Standard references on technology adoption and network externalities include Farrell and Saloner (1985) and Katz and Shapiro (1986). David's (1985) discussion of how the QWERTY keyboard became the standard for typewriters is one of the best known examples. Arthur (1989), using mathematical models which fall within the class of tools which are conventionally used in interactions based models, showed how when adoption decisions are made sequentially path dependence in technology choice

may occur, which allows inferior technologies to become locked-in. An and Kiefer (1995) show how similar results can occur through local interactions. Goolsbee and Klenow (1998) have provided evidence of the role of interaction effects in home computer adoption.

## 4. Preferences

A number of authors have used interactions-based approaches to study interdependent preferences. Föllmer (1974), in what appears to be the first explicit use of statistical mechanics methods in economics, studied an economy in which the probability that a given individual has one of two utility functions depends on the utility function of his neighbors. His work demonstrated how interactions can lead to breakdowns of the law of large numbers in large economies. Conlisk (1976) showed how to develop Markov chain models in which the distributions of behaviors at $t-1$ determined transition probabilities at $t$ and thereby are capable of producing fads in demand; Granovetter and Soong (1986) develop similar results using different methods. Bell (1995) analyzed a model in which preferences depend on the observed consumption of neighbors. Her work showed how supply effects, in which higher consumption of a commodity by others raises the price of a good for an individual, can be combined with conformity effects, in which higher consumption by others shifts the preferences of an individual toward that commodity, to produce interesting aggregate price dynamics. Darrough, Pollak, and Wales (1983), Alessie and Kapetyn (1991), Kapetyn, van de Geer, van de Stadt and Wansbeek (1997), and Binder and Pesaran (1998b) provide empirical evidence of interaction effects in consumer expenditures using a variety of modelling approaches; Andreoni and Scholz (1998) illustrate similar effects in the context of charitable contributions.

In a complementary line of work, recent authors have considered the implications of concern over relative social position on behavior, an idea whose antecedent is Duesenberry (1949) and which is explored along many dimensions in

Frank (1985). Recent important contributions include Cole, Mailath, and Postlewaite (1992) who show how relative status concerns can provide a theory of growth and Clark and Oswald (1996) who show how such concerns affect the relationship between income and well-being, and Clark and Oswald (1998) who characterize the relationship between relative status concerns and emulative behavior. Postlewaite (1997) provides an overview of the relationship between the incorporation of relative status in utility and economic theory.

**5. Behavior of political parties**

Interactions-based methods have recently proven useful in the study of political parties. In a series of papers, Kollman, Miller, and Page (1992,1997a,b) have examined the ways in which political parties evolve in response to voter preferences when there are multiple issues of concern. Their modelling typically considers how a political party will adjust its platform in response to the preferences of voters and a consideration of the behavior of the opposing party. This work has illustrated how the convergence of party platforms to a stable configuration depends sensitively on the distribution of voter preferences as well as the degree of foresight of the parties themselves.

**6. Social pathologies**

There exists evidence that a number of types of behavior which society regards as undesirable (pathological) are sustained by interaction effects. One example of this is cigarette smoking. A number of studies (Bauman and Fisher (1986), Krosnick and Judd (1982), Jones (1994)) have directly documented a role for friend and peer group behavior in predicting individual smoking probabilities. Further, well documented differences in smoking rates between black and white teenagers and between men and women within those groupings are highly

suggestive of interactions effects. Examples which are closer to the traditional concerns of economists include crime, labor market participation, out-of-wedlock births, and school attendance. Recent theoretical models of interactions and social pathologies include Akerlof and Yellen (1994), Brock and Durlauf (1995), Nechyba (1996), Lindbeck, Nyberg and Weibull (1999), Sah (1991) and Verbrugge (1999). Statistical evidence of these effects has been found in studies such as Crane (1990), Glaeser, Sacerdote, and Scheinkman (1996), Sampson, Raudenbusch, and Earls (1997) and Sucoff and Upchurch (1998); although see Gottfredson and Hirschi (1990) and Sampson and Laub (1993) for skepticism concerning the role of peer group effects with respect to the case of juvenile delinquency. Ethnographic evidence of such interactions may be found in Anderson (1990) and Duneier and Molotch (1999). Finally, Akerlof and Kranton (1998) develop a framework for understanding the psychological bases which lead to memberships in particular reference groups with attendant behavioral implications.

### 7. Information cascades

A number of authors have considered the implications of information aggregation and behavior when agents possess idiosyncratic knowledge and are attempting to learn more by observing the behavior of others. Banerjee (1992) and Bikhchandani, Hirshleifer, and Welsh (1992) have shown how such behavior can lead to informational cascades and conformity in group behavior. Caplin and Leahy (1994) show how this idea can lead to phenomena such as bank runs; Romer (1993) develops similar results in the context of asset price movements.

### 8. Evolution of science

Since Kuhn's (1970) analysis of scientific paradigms and the nature of scientific revolutions, philosophers of science have grappled with the question of

how (and in some cases whether) a community of scientists whose members are subject to conformity effects and whose objectives include non-epistemic factors such as professional status as well as epistemic factors such as better predictability succeeds in shedding scientifically inferior theories for superior ones. Recent work, best exemplified by Kitcher (1993) has explicitly modelled scientific communities as collections of interdependent researchers. This work has led authors such as Dasgupta and David (1994), David (1998), Oomes (1998) and especially Brock and Durlauf (1998) to consider formal interactions models of scientific theory choice. Using interactions-based methods, Brock and Durlauf were able to provide conditions under which scientific evidence will outweigh non-epistemic motivations and thereby provide a model of scientific progress which takes into account critiques of various social constructivists.

This chapter is designed to describe a range of methods to study interactions effects. While the interactions-based models are now fairly well developed from the perspective of theory (see Blume and Durlauf (1998a) for discussion), the econometrics literature is still in its infancy. Most of the existing econometric work has focused on the identification issues which arise for interactions-based models. The pioneering work in this regard is Manski (1993a,b,1995,1997); see as well recent surveys by Moffitt (1998) and Duncan and Raudenbusch (1998). Even here, there is substantial work which remains to be done in terms of the analysis of nonlinear as opposed to linear models. A major purpose of this chapter will be to explore identification as well as estimation in the context of structural models of interactions.

In order to facilitate this overview, we will focus on a particular class of interactions-based models, namely binary choice models with interactions. This framework has been exploited by a number of authors, including Blume (1993, 1995), Brock (1993), Brock and Hommes (1998), Durlauf (1993,1997), and Glaeser, Sacerdote, and Scheinkman (1996). The specific framework we employ is adopted primarily from Brock and Durlauf (1995). Its important advantage, from our

perspective, is that for this class of models a tight link exists between the theoretical formulation of various socioeconomic environments and the econometric analysis of those formulations.[1]

## 2. Binary choice with social interactions

### *i.* general framework

In this section, we present a baseline model of interactions. The model is capable, for particular restrictions on its parameters, of encompassing many of the theoretical treatments of social interactions which have been developed. An additional purpose of this approach is to show how these models can be analyzed using natural extensions of standard economic reasoning. Finally, as initially recognized by Blume (1993) and Brock (1993), the model is mathematically equivalent to logistic models of discrete choice. This equivalence will allow us to analyze theoretical and econometric aspects of interactions in a common framework.

We consider a population of $I$ individuals each of which faces a binary choice. These choices are denoted by an indicator variable $\omega_i$ which has support $\{-1,1\}$. Each individual makes a choice in order to maximize a payoff function $V$. In the standard binary choice formulation of economics, this payoff function is of course assumed to depend on the characteristics of the individual in question. These characteristics, in turn, are assumed to be divided into an observable (to the modeller) vector $\underset{\sim}{Z}_i$ and a pair of unobservable (to the modeller, but observable to agent $i$) random shocks, $\epsilon_i(1)$ and $\epsilon_i(-1)$. The observable vector

[1]We will not discuss the branch of the interactions literature which uses computer simulation methods to study various environments. Epstein and Axtell (1996) represent the most ambitious and wide ranging effort yet undertaken in this regard. See also Axtell, Axelrod, Epstein, and Cohen (1996) for an analysis of how to assess simulations of this type.

can include elements such as family background, role model or peer group characteristics, and past behavior. The shocks $\epsilon_i(1)$ and $\epsilon_i(-1)$ are distinct as various types of unobservable idiosyncrasies are only relevant for one of the choices. For example for the binary choice of whether to remain enrolled or dropout of school, $\epsilon_i(1)$ might refer to a shock which measures unobserved academic ability and so is only relevant if the person stays in school. Algebraically, the individual choices represent the solutions to

$$max_{\omega_i \in \{-1,1\}} V(\omega_i, \underset{\sim}{Z}_i, \epsilon_i(\omega_i)) \tag{1}$$

The standard approach to characterizing the behavior of the population of choices, an approach which renders the model econometrically estimable, is to make some assumption concerning the distribution of the $\epsilon_i(\omega_i)$'s. One common assumption is that the unobservables are independent and extreme value distributed both within and across individuals. This will imply that for a given individual, the difference between the unobservable components is logistically distributed,

$$\mu(\epsilon_i(-1) - \epsilon_i(1) \leq z) = \frac{1}{1 + exp(-\beta_i z)}; \ \beta_i \geq 0. \tag{2}$$

We use $\mu(\cdot)$ to denote probability measures throughout. The subscript $i$ here and elsewhere will be used to capture dependence on $\underset{\sim}{Z}_i$ so that, for example, $\beta_i = \beta(\underset{\sim}{Z}_i)$.

The interactions-based approach to binary choice, at least qualitatively, is based upon studying this same model once explicit attention has been given to the influence of the expected behavior of others on each individual's choice. Algebraically, each choice is described by

$$max_{\omega_i \in \{-1,1\}} V(\omega_i, \underset{\sim}{Z}_i, \mu_i^e(\underset{\sim}{\omega}_{-i}), \epsilon_i(\omega_i)) \tag{3}$$

where $\underset{\sim}{\omega}_{-i} = (\omega_1,....\omega_{i-1},\omega_{i+1},...\omega_I)$ denotes the vector of choices other than that of $i$ and $\mu_i^e(\underset{\sim}{\omega}_{-i})$ denotes that individual's beliefs concerning the choices of other agents. The nature of these beliefs, whether they are rational, etc., will be specified below. However, we will assume that beliefs are independent of the realization of any of the $\epsilon_i(\omega_i)$'s.

At this level of generality, there is of course little that can be said about the properties of the population as a whole. Hence, we make two parametric assumptions which will elucidate both basic ideas and will encompass (as special cases) a number of models which have appeared in the literature. First, we assume that the payoff function $V$ can be additively decomposed into three terms.

$$V(\omega_i,\underset{\sim}{Z}_i,\mu_i^e(\underset{\sim}{\omega}_{-i}),\epsilon_i(\omega_i)) = u(\omega_i,\underset{\sim}{Z}_i) + S(\omega_i,\underset{\sim}{Z}_i,\mu_i^e(\underset{\sim}{\omega}_{-i})) + \epsilon_i(\omega_i) \tag{4}$$

Here $u(\omega_i,\underset{\sim}{Z}_i)$ represents deterministic private utility, $S(\omega_i,\underset{\sim}{Z}_i,\mu_i^e(\underset{\sim}{\omega}_{-i}))$ represents deterministic social utility, and $\epsilon(\omega_i)$ represents random private utility. The two private utility components are standard in the econometric formulations of discrete choice. The essential difference between recent theoretical work and previous approaches to studying binary choices is the introduction of social utility considerations.

Second, we assume that this social utility term embodies a generalized quadratic conformity effect, i.e.

$$S(\omega_i,\underset{\sim}{Z}_i,\mu_i^e(\underset{\sim}{\omega}_{-i})) = -\mathrm{E}_i \sum_{j \neq i} \frac{J_{i,j}}{2} (\omega_i - \omega_j)^2 \tag{5}$$

The term $\frac{J_{i,j}}{2}$ represents the interaction weight which relates $i$'s choice to $j$'s choice and is typically assumed to be nonnegative in theoretical models, although there is no need to do so. We also treat the $J_{i,j}$ parameters as fixed; see Ioannides (1990,1997a), Kirman (1983), and Kirman, Oddou, and Weber (1986) for analyses where such parameters are stochastic using techniques from random graph theory. One can allow the $J_{i,j}$'s to depend on the characteristics of agent $j$

as well as agent $i$, so that $J_{i,j} = J(\underset{\sim}{Z}_i, \underset{\sim}{Z}_j)$.

These assumptions are sufficient to characterize the distribution of aggregate choices as a function of the distribution of various microeconomic characteristics. As a preliminary, we make two algebraic manipulations. Observe first that we can, without loss of generality, replace the private deterministic utility function of each individual with a linear function,

$$u(\omega_i, \underset{\sim}{Z}_i) = h_i \omega_i + k_i \tag{6}$$

where $h_i = h(\underset{\sim}{Z}_i)$ and $k_i = k(\underset{\sim}{Z}_i)$ are chosen so that

$$h_i + k_i = u(1, \underset{\sim}{Z}_i) \tag{7}$$

and

$$-h_i + k_i = u(-1, \underset{\sim}{Z}_i) \tag{8}$$

This linearization is permissible since the new function coincides with the original utility function on the support of the individual choices. Hence, it does not readily generalize when more than two choices are available.

Second, we expand the social utility term (5), using $\omega_i^2 = \omega_j^2 = 1$, in that

$$S(\omega_i, \underset{\sim}{Z}_i, \mu_i^e(\underset{\sim}{\omega}_{-i})) = \sum_{j \neq i} J_{i,j}(\omega_i \mathrm{E}_i(\omega_j) - 1). \tag{9}$$

which makes clear the role of pairwise interactions between each individual choice and the expected choices of others. Notice that the $J_{i,j}$ is equal to the cross-partial derivative of the social utility function, in that

$$J_{i,j} = \frac{\partial^2 V(\omega_i, \underset{\sim}{Z}_i, \mu_i^e(\underset{\sim}{\omega}_{-i}))}{\partial \omega_i \partial \mathrm{E}_i(\omega_j)} = \frac{\partial^2 S(\omega_i, \underset{\sim}{Z}_i, \mu_i^e(\underset{\sim}{\omega}_{-i}))}{\partial \omega_i \partial \mathrm{E}_i(\omega_j)}. \tag{10}$$

which means that the function measures the strategic complementarity between individual choices and the expected choices of others. See Cooper and John (1988) for a general analysis of complementarities which provides many insights which will reappear in our framework. Unlike the standard formulation of complementarities, our interactions are driven by expectations of the behavior of others, rather than by their actual behavior.

The probability that individual $i$ makes choice $\omega_i$ is equal to the probability that the utility of the choice exceeds that of $-\omega_i$,

$$\mu(\omega_i \,|\, \underset{\sim}{Z}_i,\, \mu_i^e(\underset{\sim}{\omega}_{-i})) \;=$$

$$\mu(V(\omega_i,\underset{\sim}{Z}_i,\mu_i^e(\underset{\sim}{\omega}_{-i}),\epsilon_i(\omega_i)) > V(-\omega_i,\underset{\sim}{Z}_i,\mu_i^e(\underset{\sim}{\omega}_{-i}),\epsilon_i(-\omega_i))) =$$

$$\mu(h_i\omega_i + \sum_{j \neq i} J_{i,j}\omega_i \mathrm{E}_i(\omega_j) + \epsilon_i(\omega_i) >$$

$$-h_i\omega_i - \sum_{j \neq i} J_{i,j}\omega_i \mathrm{E}_i(\omega_j) + \epsilon_i(-\omega_i)) \tag{11}$$

Letting " $\sim$ " denote "is proportional to," the logistic specification of the random utility terms means that this probability has the feature that

$$\mu(\omega_i \,|\, \underset{\sim}{Z}_i,\, \mu_i^e(\underset{\sim}{\omega}_{-i})) \sim$$

$$exp(\beta_i h_i \omega_i + \sum_{j \neq i} \beta_i J_{i,j}\omega_i \mathrm{E}_i(\omega_j)). \tag{12}$$

Since the random utility terms are independent across individuals, it must be the case that the joint set of choices obeys

$$\mu(\underset{\sim}{\omega} \;|\, \underset{\sim}{Z}_1,\ldots,\underset{\sim}{Z}_I,\mu_1^e(\underset{\sim}{\omega}_{-1}),\ldots,\mu_I^e(\underset{\sim}{\omega}_{-I})) = \prod_i \mu(\omega_i \,|\, \underset{\sim}{Z}_i,\, \mu_i^e(\underset{\sim}{\omega}_{-i})) \sim$$

$$\prod_i exp(\beta_i h_i \omega_i + \sum_{j \neq i} \beta_i J_{i,j}\omega_i \mathrm{E}_i(\omega_j)) \tag{13}$$

Equation (13) provides a general form for the joint probability measure of individual choices. It has the general form of a Gibbs measure, which is not coincidental. An important theorem in the statistical mechanics literature, due to Averintsev (1970) and Spitzer (1971a), states that models of stochastic interactions of the type which have been outlined will generically possess probability measures with Gibbs representations.

To close the model, it is necessary to specify how expectations are determined. A natural special case of this model occurs when the agents all possess rational expectations, i.e.

$$\mathrm{E}_i(\omega_j) = \mathrm{E}(\omega_j \mid \underset{\sim}{Z}_1 \ldots \underset{\sim}{Z}_I, \mathrm{E}_k(\omega_l),\ k = 1 \ldots I,\ l = 1 \ldots I) \tag{14}$$

The expectation operator on the right hand side is the mathematical expectation given by the equilibrium probability measure (13), when these same mathematical expectations are also the subjective expectations of each of the individual agents. This means that the expected values of each of the choices is constrained by a set of self-consistency conditions. In particular, the expected value of each of the individual choices for any set of beliefs will equal

$$\mathrm{E}(\omega_i) = tanh(\beta_i h_i + \sum_{j \neq i} \beta_i J_{i,j} \mathrm{E}_i(\omega_j)) \tag{15}$$

and so rational expectations require that we replace the subjective expectations with their mathematical counterparts, i.e.

$$\mathrm{E}(\omega_i) = tanh(\beta_i h_i + \sum_{j \neq i} \beta_i J_{i,j} \mathrm{E}(\omega_j)) \tag{16}$$

These equations represent a continuous mapping of $[-1,1]^I$ to $[-1,1]^I$. Therefore, it is immediate from Brouwer's fixed point theorem that there is at least one fixed point solution, which implies Theorem 1.

**Theorem 1. Existence of self-consistent equilibrium**

There exists at least one set of self-consistent expectations consistent with the binary choice model with interactions as specified by eqs. (2), (4) and (5).

By choosing particular specifications for the distribution of $Z_i$ one can generate many of the models of binary choices with interactions which have appeared in the literature. Perhaps more important, these particular specifications illustrate the interesting aggregate properties of environments with interdependent decisionmaking.

We now consider some particular $J_{i,j}$ structures in order to develop more precise properties of the population's probabilistic behavior. Page (1997) provides a valuable analysis of the role of different interaction structures in generating different aggregate properties which supplements this discussion.

*ii.* **global interactions**

One version of the binary choice model assumes that interactions across individuals are global, in the sense that each individual assigns an identical weight to the expected choice of every other member of the population. Since a person always conforms to his own behavior, this is equivalent in terms of predicted behavior to assuming that an individual assigns a common weight to all persons including himself (Brock and Durlauf (1995)) and so we assume this for expositional purposes. Formally, given $i$

$$J_{i,j} = \frac{J_i}{I} \quad \forall \ j. \tag{17}$$

Notice that we normalize the global interaction term $J_i$ by the population size $I$ for analytical convenience. This specification seems especially plausible when individual groups are determined by large aggregates such as ethnicity, religion, or region.

Global interactions imply that an individual's choice is, outside of individual-specific characteristics, only influenced by his expectation of the average choice in the population, since (17) implies that the social utility term may be rewritten (after taking the square) as

$$S(\omega_i, \underset{\sim}{Z}_i, \mu_i^e(\underset{\sim}{\omega}_{-i})) = J_i(\omega_i \mathrm{E}_i(\bar{\omega}_I) - 1) \tag{18}$$

where $\mathrm{E}_i(\bar{\omega}_I)$ denotes the subjective expectation of agent $i$ of the population average $\bar{\omega}_I$. The joint probability measure for this case equals

$$\mu(\underset{\sim}{\omega} \mid \underset{\sim}{Z}_1,\ldots,\underset{\sim}{Z}_I,\mu_1^e(\underset{\sim}{\omega}_{-1}),\ldots,\mu_I^e(\underset{\sim}{\omega}_{-I})) \sim \prod_i exp(\beta_i h_i \omega_i + \beta_i J_i \omega_i \mathrm{E}_i(\bar{\omega}_I)) \tag{19}$$

As for the general case, self-consistency requires that each individual's subjective belief concerning the average choice equals the mathematical expectation of the average choice,

$$\mathrm{E}_i(\bar{\omega}_I) = \mathrm{E}(\bar{\omega}_I \mid \underset{\sim}{Z}_1,\ldots,\underset{\sim}{Z}_I,\mu_1^e(\underset{\sim}{\omega}_{-1}),\ldots,\mu_I^e(\underset{\sim}{\omega}_{-I})) \quad \forall\, i, \tag{20}$$

which combined with the expected value of each choice, eq. (15), means that for the global interactions model, any $m$ is a self-consistent solution for the expected average choice level if it solves

$$m = \int tanh(\beta(\underset{\sim}{Z})h(\underset{\sim}{Z}) + \beta(\underset{\sim}{Z})J(\underset{\sim}{Z})m)dF_{\underset{\sim}{Z}} \tag{21}$$

where $dF_{\underset{\sim}{Z}}$ denotes the empirical probability distribution of the observable individual characteristics. When each individual possesses identical observable

characteristics, $h_i$, $\beta_i$ and $J_i$ are constant across the population, which implies that this integral reduces to the equation

$$m = tanh(\beta h + \beta J m). \tag{22}$$

This equation is easily analyzed and illustrates how multiple equilibria can emerge in interactions-based systems. Following the analysis in Brock and Durlauf (1995), these multiple equilibria can be described by Theorem 2.

**Theorem 2. Number of equilibria in the binary choice model with interactions**

*i.* If $\beta J > 1$ and $h = 0$, there exist three different values of $m$ which solve eq. (22). One of these roots is positive, one root is zero, and one root is negative.

*ii.* If $\beta J > 1$ and $h \neq 0$, there exists a threshold $H$, (which depends on $\beta$ and $J$) such that

*a.* for $|\beta h| < H$, there exist three solutions $m$ to eq. (22), one of which has the same sign as $h$, and the others possessing the opposite sign.

*b.* for $|\beta h| > H$, there exists a unique solution $m$ to eq. (22) with the same sign as $h$.

This theorem can be extended to the more general specification

$$m = \int tanh(\beta h(\underset{\sim}{Z}) + \beta J m) dF_{\underset{\sim}{Z}} \tag{23}$$

which differs from eq. (21) in that here $\beta$ and $J$ are assumed to be constant across

individuals. The case of heterogeneous $h_i$'s is of particular interest when considering the econometric implementation of the model.

In order to generalize our theorem, we define the function $R(\cdot)$ by

$$R(m) = \int tanh(\beta h(\underset{\sim}{Z}) + \beta J m) dF_{\underset{\sim}{Z}} \tag{24}$$

so that the integral can be treated as a function of $m$. Suppose that $dF_{\underset{\sim}{Z}}$ is symmetrically distributed with mean 0 and variance $s$ and that $h(\underset{\sim}{Z})$ is symmetric about the origin. This implies that $R(0) = 0$ given $tanh(-x) = -tanh(x)$ and the assumed symmetry in $h$ and $dF_{\underset{\sim}{Z}}$. Next, define

$$r(m) = \int tanh'(\beta h(\underset{\sim}{Z}) + \beta J m) dF_{\underset{\sim}{Z}} \tag{25}$$

Observe that

$$R'(0) = \beta J \int tanh'(\beta h(\underset{\sim}{Z})) dF_{\underset{\sim}{Z}} = \beta J r(0) > 0. \tag{26}$$

This means that for sufficiently small $\beta J$, $\beta J r(0) < 1$ but if $\beta J > r(0)^{-1}$ then $R'(0) > 1$ and hence at least two new equilibria exist besides $m = 0$. On the other hand, note that for any pair $m_1$ and $m_2$

$$| R(m_1) - R(m_2) | \leq \beta J | m_1 - m_2 | \tag{27}$$

using (23), the mean value theorem, and the facts that the *tanh* function is bounded between $-1$ and 1 and $dF_{\underset{\sim}{Z}}$ is a probability measure. If $\beta J < 1$, then this is a contraction mapping and there exists only one solution to eq. (23) in this case. Hence the $m = 0$ solution bifurcates into at least three solutions as $\beta J$ increases beyond 1. Notice that unlike the case of homogeneous $h$'s, we have not ruled out the possibility that more than three equilibria exist. We summarize this as a corollary.

**Corollary 1. Number of equilibria in binary choice model with global interactions and individual heterogeneity**

If $h(\underset{\sim}{Z})$ is distributed symmetrically about the origin, then

*i.* If $\beta J < 1$, then the self-consistent equilibrium in eq. (23) is unique.

*ii.* If $\beta J > 1$, then there exist at least three self-consistent solutions to eq. (23).

***iii.* local interactions**

Local interactions models typically assume that each agent interacts directly with only a finite number of others in the population. For each $i$, the set of $j$'s with whom he has interactions is referred to as his neighborhood and is denoted by $n_i$. While residential neighborhoods have been a longstanding focus of the interactions literature, the models we analyze have much broader applicability.

In a local interactions model, the notion of neighborhood-level interactions is captured by a restriction on the interaction weights $J_{i,j}$ of the general form

$$J_{i,j} = 0 \text{ if } j \notin n_i \tag{28}$$

Of course, the global interactions model can be treated as a special case of a neighborhoods model, one where all other members of the population are members of each $i$'s neighborhood.

Depending on the application, the index $i$ has been interpreted differently. For example, in Föllmer (1974) or Glaeser, Sacerdote, and Scheinkman (1995), $|i-j|$ measures the distance between individuals, whereas it is treated as an

index of technological similarity as in Durlauf (1993). This allows one to construct a neighborhood for agent $i$ by taking all agents within some fixed distance from $i$. (The distance can vary with direction.) This latter assumption is the source of the term "local." For purposes of analysis of finite systems, it is typical to locate actors on a torus so that distance can be defined symmetrically for all agents. (A two dimensional torus is formed out of a $k \times k$ lattice by connecting the east/west and north/south boundaries so as to ensure that each element of the resulting system has four nearest neighbors.) For agents located on a torus, one can rewrite social utility as

$$S(\omega_i, \underset{\sim}{Z}_i, \mu_i^e(\underset{\sim}{\omega}_{-i})) = -\mathrm{E}_i \sum_{j \in n_i} \frac{J_{i,j}}{2} (\omega_i - \omega_j)^2 \tag{29}$$

with associated individual probability measure

$$\mu\ (\omega_i \mid \underset{\sim}{Z}_i,\ \mu_i^e(\underset{\sim}{\omega}_{-i})) \sim$$

$$exp(\beta_i h_i \omega_i + \sum_{j \in n_i} \beta_i J_{i,j} \omega_i \mathrm{E}_i(\omega_j)). \tag{30}$$

and joint probability measure

$$\mu(\underset{\sim}{\omega}\ \mid \underset{\sim}{Z}_1,\ldots,\underset{\sim}{Z}_I, \mu_1^e(\underset{\sim}{\omega}_{-1}),\ldots,\mu_I^e(\underset{\sim}{\omega}_{-I})) = \prod_i \mu(\omega_i \mid \underset{\sim}{Z}_i,\ \mu_i^e(\underset{\sim}{\omega}_{-i})) \sim$$

$$\prod_i exp(\beta_i h_i \omega_i + \sum_{j \in n_i} \beta_i J_{i,j} \omega_i \mathrm{E}_i(\omega_j)) \tag{31}$$

A special case of the local interactions model occurs when local interactions are homogeneous, which means 1) all neighborhoods have the same size which we denote $N$ and 2) within a neighborhood, all interaction weights are equal to a common $J$. In this special case, social utility will equal

$$S(\omega_i, \underset{\sim}{Z}_i, \mu_i^e(\underset{\sim}{\omega}_{-i})) = J\omega_i \sum_{j \in n_i} \mathrm{E}_i(\omega_j) \tag{32}$$

where $N$ denotes the number of members of a neighborhood. Under rational expectations it is immediate that one joint probability measure for agents' choices is

$$\mu(\underset{\sim}{\omega}) \sim exp(\beta h \sum_i \omega_i + \beta N J \sum_i \omega_i E(\omega)) \tag{33}$$

where

$$\mathrm{E}(\omega) = tanh(\beta h + \beta N J \mathrm{E}(\omega)) = \mathrm{E}(\omega_i) \ \forall i. \tag{34}$$

which implies the following theorem.

**Theorem 3. Relationship between global and local interactions models**

Any equilibrium expected individual and average choice level $m$ for the global interactions model is also an equilibrium expected individual and average choice in a homogeneous local interactions model.

This result might initially appear odd, given the explicit local interaction structure of preferences. In fact, the equivalence is not surprising. When all expectations are identical, and the sample mean is required to equal the population mean, then agents are all implicitly connected to one another through the expectations formation process. To be clear, the local interactions model can exhibit equilibria which are different from that of the global case.

Focus on the case where each individual is required to possess identical $\mathrm{E}(\omega_i)$'s is not required by the logic of the local interactions model. There has been little work on the existence and characterization of asymmetric equilibrium $\mathrm{E}(\omega_i)$'s, i.e. equilibria where the expected values differ across agents. Examples of asymmetric equilibria of this type may be found in Blume and Durlauf (1998b).

A trivial example can be produced by taking two environments which exhibit global interactions and multiple equilibria and defining them as a common population.

Finally, it is worth noting that when interactions between decisions are all intertemporal, then the assumption of extreme-valued random utility increments can be dropped. The equilibrium properties of the dynamic models in this section can be recomputed under alternative probability densities such as probit which are popular in the discrete choice work. In fact, under the mean field analysis of global interactions, alternative specifications incorporate probit or other densities as well. In both cases, the large scale properties of models under alternative error distributions are largely unknown.

***iv.*** **relationship to statistical mechanics**

The models we have thus far outlined bear a close relationship to models in statistical mechanics. A standard question in statistical mechanics concerns how a magnet can exist in nature. A magnet is defined as a piece of iron in which a majority of the atoms are either spinning up or down. Since there is no physical reason why atoms should be more likely to spin up or down when considered in isolation, the existence of a natural magnet, which requires literally billions of atoms to be polarized towards one type of spin, would seem extraordinary unlikely by the law of large numbers. As a result, statistical mechanics models are based on the primitive idea that the probability that one atom has a given spin is an increasing function of the number of atoms with the same spin within the atom's neighborhood. For the Ising model of ferromagnetism, the assumption is that atoms are arrayed on a $2-$ (or higher) dimensional integer lattice, so that

$$\mu(\omega_i \mid \textit{spins of all other atoms in material}) =$$

$$\mu(\omega_i \mid \omega_j \text{ such that } \mid i-j \mid = 1) \sim exp(\beta J \omega_i \sum_{\mid i-j \mid = 1} \omega_j) \tag{35}$$

For the Curie – Weiss model, the physical interaction structure is assumed to be such that each atom's spin is probabilistically dependent on the average spin in the system, so that

$$\mu(\omega_i \mid \textit{spins of all other atoms in material}) \sim exp(\beta J \omega_i \overline{\omega}) \tag{36}$$

Hence our models of binary choice with social interactions are mathematically quite similar to physical models of magnetism.

An important difference, however, does exist. While our socioeconomic model embeds pairwise interactions via the products of individual choices $\omega_i$ with the expected choices of others, the physical models are based upon conditional probabilities which depend on the products of the realized individuals choices for all pairs of individuals. Interestingly, the physics literature has also dealt with expectations-based interactions. It turns out that models with interactions across realizations are extremely difficult to analyze, so physicists have developed what is referred to a "mean-field approximation" to various ferromagnetism models. A mean field approximation amounts to replacing certain terms in an original model with their mathematical expectation. Hence, the mean field approximation for the conditional probability of the spin of a given atom for the Curie-Weiss model is

$$\mu(\omega_i) \sim exp(\beta J \omega_i \mathrm{E}(\overline{\omega}_I)) \tag{37}$$

which is of the same form as eq. (12) when agents possess identical $\underset{\sim}{Z}_i$'s and $J_{i,j} = \frac{J}{I}$. Of course, what is an approximate model in a physical context is an exact model in the socioeconomic context we have been analyzing, at least given our behavior primitives. This difference occurs because our behavior assumption is that individuals interact through their expectation's of one another's behavior,

rather than through realizations.

This last remark relates to a more general consideration in the use of statistical mechanics methods by social scientists. A basic conceptual difference exists between social and physical environments which contain interactions. Physical (and many) mathematical models of interactions typically take as primitives the conditional probabilities linking elements of a system, i.e $\mu(\omega_1 \mid \underset{\sim}{\omega}_{-1}),\ldots,\mu(\omega_I \mid \underset{\sim}{\omega}_{-I})$. Analysis of the model considers the existence and (if so) properties of whatever joint probability measures are consistent with the conditional ones. In socioeconomic contexts, it is more natural to take preferences, beliefs, and technologies as primitives and from them determine what conditional probability relationships will hold. Hence, statistical mechanics and related models cannot be employed in socioeconomic contexts without determining what socioeconomic primitives will lead to a particular conditional probability representation. Further, the purposefulness of the objects of analysis in social science contexts also means that issues of the endogeneity of neighborhoods and the potential for the existence of institutions which coordinate collective action will naturally arise. These issues have no analog in physical contexts and are suggestive of the limitations in importing methods from physics into socioeconomic studies.

### *v.* social planning problem

Our analysis thus far has assumed that individual decisions are not coordinated. An alternative approach is to examine how decisions would be made when coordinated by a social planner. Beyond its use in developing welfare comparisons and other contrast with the noncooperative case, the social planner's solution may have empirical content in some contexts. As described by Coleman (1988, 1990 chapter 12), the evolution of social capital, defined to include aspects of social structure which facilitate coordination across individuals and which may

be embedded either in personal mores or organizations such as churches or schools, implies that in many types of social situations, coordinated behavior can emerge.

In order to do this, it is necessary to be more precise in the formulation of the underlying game played by members of the population. As before, we consider a population of $I$ individuals each with payoff function $V(\omega_i,\underset{\sim}{Z}_i,\mu_i^e(\underset{\sim}{\omega}_{-i}),\epsilon_i(\omega_i))$. The random functions $\epsilon_i(\cdot)$ are assumed to be observed by the members of the population, so that each agent $i$ knows the realizations of $\epsilon_j(\cdot)\ \forall\ j \neq i$. We further assume that the distribution of these random components is described by eq. (2). Hence in terms of timing, nature draws the random functions $\epsilon_i(\cdot)$ and reveals them to the entire population. Second, players play the game $G$ defined by

$$G = \{V(\omega_i,\underset{\sim}{Z}_i,\mu_i^e(\underset{\sim}{\omega}_{-i}),\epsilon_i(\omega_i)),\ i = 1...I\} \tag{38}$$

where $\mu_i^e(\underset{\sim}{\omega}_{-i})$ denotes their beliefs about the behavior of other agents and is conditioned on nature's draw of the random functions.

With respect to this environment, an obvious benchmark is a perfect foresight Nash equilibrium. By this, we mean that each player knows the $\epsilon_i(\cdot)$ functions for every agent and forms beliefs about the resultant choices in the population $\mu_i^e(\underset{\sim}{\omega}_{-i})$ which are confirmed in equilibrium. If each player is playing a pure strategy, this means that $\mu_i^e(\underset{\sim}{\omega}_{-i}) = \underset{\sim}{\omega}_{-i}$ so that a perfect foresight pure strategy equilibrium is a set of choices $\underset{\sim}{\omega}$ such that for all $i$

$$\omega_i = argmax_{\gamma \in \{-1,1\}} V(\gamma,\underset{\sim}{Z}_i,\underset{\sim}{\omega}_{-i},\epsilon_i(\gamma)). \tag{39}$$

For the analogous mixed strategy equilibrium, let $\underset{\sim}{\pi}_i = (\pi_{i,-1},\pi_{i,1})$ denote the row vector of probability weights assigned by agent $i$ to the two choices. Then $\underset{\sim}{\Pi}_i = (\underset{\sim}{\pi}_1 ... \underset{\sim}{\pi}_I)$ denotes a perfect foresight Nash equilibrium if each $\underset{\sim}{\pi}_i$ is consistent with

$$\underset{\sim}{\pi}_i = argmax_{\underset{\sim}{\gamma}_i}$$

$$\gamma_{i,1} V(1, \underset{\sim}{Z}_i, \underset{\sim}{\Pi}_{-i}, \epsilon_i(1) + \gamma_{i,-1} V(-1, \underset{\sim}{Z}_i, \underset{\sim}{\Pi}_{-i}, \epsilon_i(-1)))$$

$$such\ that\ \gamma_{i,-1}, \gamma_{i,1} \geq 0\ and\ \ \gamma_{i,-1} + \gamma_{i,1} = 1, \tag{40}$$

where $E_{\underset{\sim}{\gamma}}$ means that agent $i$ plays the mixture $\underset{\sim}{\gamma}_i$ against the mixtures played by the other agents, $\underset{\sim}{\Pi}_{-i} = (\underset{\sim}{\pi}_1, \cdots, \underset{\sim}{\pi}_{i-1}, \underset{\sim}{\pi}_{i+1}, \cdots, \underset{\sim}{\pi}_I)$. Mixture $\underset{\sim}{\gamma}_i$ means that $i$ chooses 1 with probability $\gamma_{i,1}$ and chooses $-1$ with probability $\gamma_{i,-1}$. It is a standard result that a mixed Nash equilibrium of this type will always exist, although a pure strategy Nash equilibrium may not.

Alternatively, a limited information Nash equilibrium can be characterized when agents make choices without knowledge of the $\epsilon_i(\cdot)$ functions for agent's other than themselves. In terms of timing, one can think of agents forming beliefs $\mu_i^e(\underset{\sim}{\omega}_{-i})$ before any $\epsilon_i(\cdot)$'s are realized, nature then drawing the $\epsilon_i(\cdot)$'s, revealing $\epsilon_i(\cdot)$ to agent $i$, and each $i$ then choosing $\omega_i$. For this case,

$$\omega_i = argmax_{\gamma \in \{-1,1\}} V(\gamma, \underset{\sim}{Z}_i, \mu_i^e(\underset{\sim}{\omega}_{-i}), \epsilon_i(\gamma)), \tag{41}$$

when $\mu_i^e(\underset{\sim}{\omega}_{-i}) = \mu(\underset{\sim}{\omega}_{-i} \,|\, \underset{\sim}{Z}_j\ \forall\ j)\ \forall\ i$, so that each agents belief's are consistent with the model. This is the equilibrium concept we have employed above.

In contrast to these noncooperative environments, we may characterize a social planner's perfect foresight problem as choosing $\underset{\sim}{\omega}$ in order to maximize total utility in the population, i.e.

$$max_{\underset{\sim}{\omega}} \sum_{i=1}^{I} V(\omega_i, \underset{\sim}{Z}_i, \mu_i^e(\underset{\sim}{\omega}_{-i}), \epsilon_i(\omega_i)) \tag{42}$$

From (42), one can in principle compute quantities such as the expected average payoff under a social planner and contrast it with their counterparts under the

two noncooperative environments.

In order to perform such a comparison, however, analytical tractability becomes a problem. To see this, notice that for our global interactions model, the social planner's problem becomes

$$max_{\underset{\sim}{\omega}} \sum_{i=1}^{I} (h_i\omega_i - \frac{J}{2}(\omega_i - \bar{\omega}_I)^2 + \epsilon_i(\omega_i)) \tag{43}$$

Unfortunately, $\sum_{i=1}^{I} \epsilon_i(\omega_i)$ is not independent and extreme value distributed over the $2^I$ possible configurations of $\underset{\sim}{\omega}$ even though the individual $\epsilon_i(\omega_i)$'s are distributed that way. One way around this problem is, following Brock and Durlauf (1995), to replace this original social planner's problem with an approximate problem

$$max_{\underset{\sim}{\omega}} \sum_{i=1}^{I} (h_i\omega_i - \frac{J}{2}(\omega_i - \bar{\omega}_I)^2 + \epsilon_i^*(\omega_i)) \tag{44}$$

where $\sum_{i=1}^{I} \epsilon_i^*(\omega_i)$ is itself extreme value distributed. One can require that the variance of the errors in the approximate social planner's problem to equal those in the original problem in order to achieve some calibration between the two problems.

This approximate social planner's problem can be analyzed by replacing $\mathrm{E}_i(\bar{\omega}_I)$ with $\bar{\omega}_I$ in eq. (19). The probability measure characterizing the joint choice of $\underset{\sim}{\omega}$ equals

$$\mu(\underset{\sim}{\omega}) = \frac{exp(\beta(\sum_{i=1}^{I} h_i\omega_i + \frac{J}{I}(\sum_{i=1}^{I} \omega_i)^2))}{\sum_{\nu_1 \in \{-1,1\}} \cdots \sum_{\nu_I \in \{-1,1\}} exp(\beta(\sum_{i=1}^{I} h_i\nu_i + \frac{J}{I}(\sum_{i=1}^{I} \nu_i)^2))}. \tag{45}$$

In order to analyze this probability measure, which is known in the statistical mechanics literature as the Curie-Weiss model, it is necessary to

eliminate the $(\sum_{i=1}^{I} \omega_i)^2$ terms in eq. (45). This calculation is complicated and may be found in the Mathematical Appendix; further analysis appears in Brock (1993). A result currently exists only for the case $h_i = h$ and only for the large economy limit; however, Amaro de Matos and Perez (1991) suggests that for the large economy limit generalization to heterogeneous $h_i$'s is possible. The appendix verifies Theorem 4.

**Theorem 4. Expected average choice under social planner for binary choice model with interactions**

Let $m^*$ denote the root of $m^* = tanh(\beta h + \beta J m^*)$ with the same sign as $h$. If eq. (39) characterizes the joint distribution of individual choices as determined by a social planner, then

$$lim_{I \Rightarrow \infty} \mathrm{E}(\overline{\omega}_I) = m^*. \tag{46}$$

One aspect of this theorem is intuitive, in that a planner would choose that average choice level in which the interaction effects and the private deterministic utility comparisons work together. What is perhaps surprising is that the social planner's equilibrium is sustainable as an equilibrium in the limited information noncooperative environment. However, this result is somewhat special to the functional form originally assumed for individual deterministic social utility. If the original social utility term had been $J\omega_i \mathrm{E}(\overline{\omega}_I)$, then the noncooperative equilibrium average choice level would be the same as for the case we have studied, but the analogous social planner's problem would choose that root of $m^* = tanh(\beta h + 2\beta J m^*)$ with the same sign as $h$ (Brock and Durlauf (1995)), which would mean it is not supportable in the limited information noncooperative environment.

*vi.* **linear-in-means model**

Much of the empirical work on interaction effects has assumed that the behavior variable $\omega_i$ has continuous support and depends linearly on various individual and neighborhood effects. These assumptions permit a researcher to use ordinary least squares methods, which will be discussed below. While these empirical papers generally do not consider what decision problems generate their econometric specifications, it is straightforward to do so. For a trivial example, suppose that an individual solves

$$max_{\omega_i \in (-\infty,\infty)} -\tfrac{1}{2}(\omega_i - \omega_i^*)^2 \tag{47}$$

where $\omega_i^*$ is a reference behavior level to which individual $i$ prefers to conform. When this reference behavior level equals $h_i + J\mathrm{E}_i(\overline{\omega}_I) + \epsilon_i$, it is immediate that

$$\omega_i = h_i + J\mathrm{E}_i(\overline{\omega}_I) + \epsilon_i. \tag{48}$$

This is the type of equation studied by Manski (1993), Moffitt (1998), Duncan and Raudenbusch (1998), among others.

## 2. Identification: Basic issues

In this section, we describe the identification of interactions-based models in cross-sections. Identification is a concern in these cases because of the likelihood that group versus individual determinants of individual behavior are likely to be correlated. Hauser (1970) provides an early and clever analysis of how these correlations can, if not properly accounted for, lead to spurious inferences. We recommend this paper as an example of how powerful intuitive reasoning (as well as good common sense) can complement and foreshadow formal analysis.

Manski (1993a,b,1997) has pioneered the study of the identification of interaction effects, and we will follow his treatment closely. In his work, Manski, distinguishes between three explanations for correlated behavior within groups:

"*endogenous effects*, wherein the propensity of an individual to behave in some way varies with the behaviour of the group...*exogenous* (*contextual*) effects, wherein the propensity of an individual to behave in some way varies with the exogenous characteristics of a group...*correlated* effects, wherein individuals in the same group tend to behave similarly because they have similar individual characteristics or face similar institutional environments" (Manski (1993a), pg. 532)

The treatment of identification problems in terms of the ability to distinguish these different effects in data seems to us very useful and so we employ it throughout.

For purposes of discussing identification and other econometric aspects of interaction-based models, we begin with a baseline set of data assumptions which will apply both to the binary choice model and to the linear-in-means model. We assume that the econometrician has available a set of observations on $I$ individuals. We assume that each individual is drawn randomly from a set of neighborhoods. Within each neighborhood, all interactions are global. For notational purposes, we denote individuals as $i$ and the neighborhood (which means the set of other individuals who influence $i$ through interactions) as $n(i)$. We assume that our original $\underset{\sim}{Z}_i$ vector can be partitioned into an $r$-length vector of individual-specific observables $\underset{\sim}{X}_i$ and an $s$-length vector of exogenously determined neighborhood observables $\underset{\sim}{Y}_{n(i)}$ associated with each individual in the sample. This will allow us to replace the private utility component $h_i$ in our theoretical discussion with a linear specification

$$h_i = k + c'\underset{\sim}{X}_i + d'\underset{\sim}{Y}_{n(i)}. \tag{49}$$

Notice that this specification means that none of the individual-specific observables $\underset{\sim}{X}_i$ or neighborhood observables $\underset{\sim}{Y}_{n(i)}$ contains a constant term. We

will maintain this assumption throughout. Within a neighborhood, all interactions are assumed to be global and symmetric, so that there is a single parameter $J$ which indexes interactions.

Recall that $m^e_{n(i)}$ is agent $i$'s subjective expectation of the average choice in neighborhood $n(i)$. In the subsequent discussion, it will be useful to distinguish between $m^e_{n(i)}$ and $m_{n(i)}$, the mathematical expectation of the average choice in a neighborhood under self-consistency. (We will specify the information sets under which self-consistency is calculated below.) The reason for this is that we will have need to distinguish between the data in a statistical exercise and the mathematical solution to a model. Of course, $m^e_{n(i)} = m_{n(i)}$ is part of our maintained assumption in the analysis, so there is no loss of generality in doing this. For purposes of discussion of identification, we are therefore either implicitly assuming that the neighborhoods are arbitrarily large, so that the neighborhood sample average can be used in place of the expected value or that accurate survey data are available. We finally assume that the errors are independent across individuals and that $\mu(\epsilon_i(\omega_i) - \epsilon_i(-\omega_i) \mid \underset{\sim}{X}_i, \underset{\sim}{Y}_{n(i)}, m^e_{n(i)}) = \mu(\epsilon_i(\omega_i) - \epsilon_i(-\omega_i))$ for the binary choice model, and $\mathrm{E}(\epsilon_i \mid \underset{\sim}{X}_i, \underset{\sim}{Y}_{n(i)}, m^e_{n(i)}) = 0$ for the linear-in-means model.

Our strategy of using individual level data has an important advantage: to the extent that the parameters of the individual model are identified, one can infer whether or not multiple equilibria exist with respect to population aggregates. This can be done without consideration of an equilibrium selection rule because population aggregates are always treated as independent variables in the analysis. Hence we can circumvent some of the problems described in Jovanovic (1989).

*i.* **binary choice**

For the binary choice model, we consider the identification based on a

naive estimator of the parameters of the model. By naive, we refer to the case where a logistic regression is computed which does not impose the relationships between neighborhood means. In this case, the conditional likelihood function for the set of individual choices will have a standard logistic form. Using our theoretical model of global interactions (and exploiting symmetry of the logistic density function), the likelihood is

$$L(\underset{\sim}{\omega}_I \mid \underset{\sim}{X}_i, \underset{\sim}{Y}_{n(i)}, m^e_{n(i)} \ \forall \ i) =$$

$$\prod_i \mu(\omega_i = 1 \mid \underset{\sim}{X}_i, \underset{\sim}{Y}_{n(i)}, m^e_{n(i)})^{\frac{1+\omega_i}{2}} \cdot \mu(\omega_i = -1 \mid \underset{\sim}{X}_i, \underset{\sim}{Y}_{n(i)}, m^e_{n(i)})^{\frac{1-\omega_i}{2}} \sim$$

$$\prod_i (exp(\beta k + \beta c' \underset{\sim}{X}_i + \beta d' \underset{\sim}{Y}_{n(i)} + \beta J m^e_{n(i)})^{\frac{1+\omega_i}{2}}$$

$$\cdot exp(-\beta k - \beta c' \underset{\sim}{X}_i - \beta d' \underset{\sim}{Y}_{n(i)} - \beta J m^e_{n(i)})^{\frac{1-\omega_i}{2}}) \tag{50}$$

As is standard for logistic models, the complete set of model parameters is not identified as $k$, $c'$, $d'$ and $J$ are each multiplied by $\beta$. We therefore proceed under the normalization $\beta = 1$.

The reason that identification is a concern in a model like this is the presence of the term $m^e_{n(i)}$ in the likelihood function. Since this terms embodies a rationality condition, it is a function of other variables in the likelihood function. Specifically, we assume that

$$m^e_{n(i)} = m_{n(i)} = \int tanh(k + c' \underset{\sim}{X} + d' \underset{\sim}{Y}_{n(i)} + J m_{n(i)}) dF_{\underset{\sim}{X} \mid \underset{\sim}{Y}_{n(i)}}. \tag{51}$$

Here $F_{\underset{\sim}{X} \mid \underset{\sim}{Y}_{n(i)}}$ denotes the conditional distribution of $\underset{\sim}{X}$ in neighborhood $n(i)$ given the neighborhood characteristics $\underset{\sim}{Y}_{n(i)}$. What this means is that each agent is assumed to form the conditional probabilities of the individual characteristics in a neighborhood given the aggregates which determine his or her payoffs. Since one

can always add elements of $\underset{\sim}{Y}_{n(i)}$ with zero coefficients to the payoff equation for agents, this is without loss of generality.

Rather than prove identification for the particular case where the theoretical model is logistic (see McFadden (1974) and Amemiya (1993, chapter 9) for proofs for this case) we prove identification for an arbitrary known distribution function for the random payoff terms. Specifically, we assume that the conditional probability of individual $i$'s choice can be written as

$$\mu(\epsilon(\omega_i) - \epsilon(-\omega_i) \leq z \mid \underset{\sim}{X}_i, \underset{\sim}{Y}_{n(i)}, m^e_{n(i)}) = F(z \mid k + c'\underset{\sim}{X}_i + d'\underset{\sim}{Y}_{n(i)} + Jm^e_{n(i)}) \quad (52)$$

where $F$ is a known probability distribution function that is continuous and strictly increasing in $z$.

We consider identification based on a naive estimator of the parameters of the model. By naive, we refer to the situation where parameter estimates for the model are computed which do not impose the rational expectations condition between neighborhood means and neighborhood characteristics, but rather uses these variables as regressors. Hence, we assume that $m^e_{n(i)}$ is known to the researcher; see discussion below for the case when $m^e_{n(i)}$ is not observable.

To formally characterize identification, we employ the following notation. Define $supp(\underset{\sim}{X}, \underset{\sim}{Y}, m^e)$ as the joint support of the distribution of $(\underset{\sim}{X}_i, \underset{\sim}{Y}_{n(i)}, m^e_{n(i)})$. Intuitively, the definition of identification we employ says that a model is identified if there do not exist two distinct sets of parameter values each of which produces (for all subsets of $\underset{\sim}{X}$ and $\underset{\sim}{Y}$ which occur with positive probability) identical probabilities for individual choices and which are also self-consistent.

**Definition. Global identification in the binary choice model with interactions and self-consistent expectations**

The binary choice model is globally identified if for all parameter pairs $(k,c,d,J)$

and $(\overline{k}\,,\overline{c}\,,\overline{d}\,,\overline{J}\,)$

$$k + c'\underset{\sim}{X}_i + d'\underset{\sim}{Y}_{n(i)} + Jm^e_{n(i)} = \overline{k}\; + \overline{c}'\underset{\sim}{X}_i + \overline{d}'\underset{\sim}{Y}_{n(i)} + \overline{J}m^e_i \tag{53}$$

and

$$m^e_{n(i)} = m_{n(i)} =$$

$$\int \omega_i dF(\omega_i \mid k + c'\underset{\sim}{X} + d'\underset{\sim}{Y}_{n(i)} + Jm_{n(i)})dF_{\underset{\sim}{X}\mid\underset{\sim}{Y}_{n(i)}} =$$

$$\int \omega_i dF(\omega_i \mid \overline{k}\; + \overline{c}'\underset{\sim}{X} + \overline{d}'\underset{\sim}{Y}_{n(i)} + \overline{J}\, m_{n(i)})dF_{\underset{\sim}{X}\mid\underset{\sim}{Y}_{n(i)}} \tag{54}$$

$\forall\ (\underset{\sim}{X}_i,\underset{\sim}{Y}_{n(i)},m^e_{n(i)}) \in supp(\underset{\sim}{X},\underset{\sim}{Y},m^e)$ imply that $(k,c,d,J) = (\overline{k}\,,\,\overline{c}\,,\overline{d}\,,\overline{J}\,)$

In order to establish conditions under which identification can hold we follow the argument in Manski (1988), Proposition 5, and state the following Proposition, whose proof appears in the Mathematical Appendix. The assumptions we make are clearly sufficient rather than necessary; weakening the assumptions is left to future work. In interpreting the assumptions, note that Assumption $i$ is the one used by Manski to identify this model when there are no endogenous effects, i.e. if $J$ is known a priori to be 0. The assumption, of course does nothing more than ensure that the individual and contextual regressors are not linearly dependent. The additional assumptions are employed to account for the fact that $m_{n(i)}$ is a nonlinear function of the contextual effects.

**Theorem 5. Sufficient conditions for identification to hold in the binary choice model with interactions and self-consistent beliefs**

Assume

*i.* $supp(\underset{\sim}{X}_i, \underset{\sim}{Y}_{n(i)})$ is not contained in a proper linear subspace of $R^{r+s}$.

*ii.* $supp(\underset{\sim}{Y}_{n(i)})$ is not contained in a proper linear subspace of $R^s$.

*iii.* No element of $\underset{\sim}{X}_i$ or $\underset{\sim}{Y}_{n(i)}$ is constant.

*iv.* There exists at least one neighborhood $n_0$ such that conditional on $\underset{\sim}{Y}_{n_0}$, $\underset{\sim}{X}_i$ is not contained in a proper linear subspace of $R^r$.

*v.* None of the regressors in $\underset{\sim}{Y}_{n(i)}$ possesses bounded support.

*vi.* $m_{n(i)}$ is not constant across all neighborhoods $n$.

Then, $(k,c,d,J)$ is identified relative to any distinct alternative $(\overline{k}, \overline{c}, \overline{d}, \overline{J})$.

*ii.* **linear-in-means model**

Identification in the binary choice model with interactions can be contrasted with the case of the analogous linear-in-means model,

$$\omega_i = k + c' \underset{\sim}{X}_i + d' \underset{\sim}{Y}_{n(i)} + J m^e_{n(i)} + \epsilon_i. \tag{55}$$

The unique self-consistent solution $m_{n(i)}$ for the linear-in-means model is easily seen, by applying an expectations operator to both sides of the individual behavioral equation, to be

$$m_{n(i)} = \frac{k + c' \mathrm{E}(\underset{\sim}{X}_i \mid \underset{\sim}{Y}_{n(i)}) + d' \underset{\sim}{Y}_{n(i)}}{1 - J}. \tag{56}$$

where $\mathrm{E}(\underset{\sim}{X}_i \mid \underset{\sim}{Y}_{n(i)})$ denotes the expected value of the individual controls given the

neighborhood characteristics. Hence, following the argument in Manski (1993), one can construct a reduced form expression for individual choices,

$$\omega_i = \frac{k}{1-J} + c' \underset{\sim}{X}_i + \frac{J}{1-J} d' \underset{\sim}{Y}_{n(i)} + \frac{J}{1-J} c' \mathrm{E}(\underset{\sim}{X}_i \,|\, \underset{\sim}{Y}_{n(i)}) + \epsilon_i. \tag{57}$$

In this equation, we have $2r+s+1$ regressors and $r+s+2$ parameters. The possibility for identification in this model therefore will depend on which, if any, of the regressors in the reduced form are linearly independent (i.e. their variance covariance matrix is of full rank). For example if $\mathrm{E}(\underset{\sim}{X}_i \,|\, \underset{\sim}{Y}_{n(i)})$ is linearly dependent on $\underset{\sim}{Y}_{n(i)}$, then it is obvious that the model parameters are not identified. More generally, it is necessary for identification that the dimension of the linear space spanned by the regressors is at least equal to the number of structural parameters, i.e. $r+s+2$; otherwise, one cannot map the reduced form coefficients back to the structural parameters. Hence, one can state the following theorem.

**Theorem 6. Necessary conditions for identification in the linear-in-means model with interactions and self-consistent beliefs**

In the linear-in-means model it is necessary for identification of the model's parameters that

*i.* The dimension of the linear space spanned by elements of $(1, \underset{\sim}{X}_i, \underset{\sim}{Y}_{n(i)})$ is $r+s+1$.

*ii.* The dimension of the linear space spanned by the elements of $(1, \underset{\sim}{X}_i, \underset{\sim}{Y}_{n(i)}, \mathrm{E}(\underset{\sim}{X}_i \,|\, \underset{\sim}{Y}_{n(i)}))$ is at least $r+s+2$.

Notice that the conditions of this theorem, while analogous to those in the theorem for identification in the binary choice model, are now necessary and not

sufficient. This is because sufficient conditions will depend on the model parameters. For example, if $c = \underset{\sim}{0}$, then the fact that $\mathrm{E}(\underset{\sim}{X}_i \mid \underset{\sim}{Y}_{n(i)})$ is linearly independent of the regressors $\underset{\sim}{X}_i$ and $\underset{\sim}{Y}_{n(i)}$ will not eliminate collinearity of $m_{n(i)}$ and $\underset{\sim}{Y}_{n(i)}$ in the structural equation (55) and hence will leave only $s+1$ regression coefficients in the reduced form available to identify $k$, $J$ and $d$, which is not enough.

This theorem is an extension of Manski's (1993) result on the nonidentifiability of contextual versus endogenous effects. Manski's analysis assumes that there is a one-to-one correspondence between the individual control variables $\underset{\sim}{X}_i$ and the neighborhood control variables $\underset{\sim}{Y}_{n(i)}$ so that for any individual-level variable that influences behavior, the neighborhood average of that variable also influences behavior. For example, if one controls for individual education, one also controls for average neighborhood education. In this case, $\mathrm{E}(\underset{\sim}{X}_i \mid \underset{\sim}{Y}_{n(i)}) = \underset{\sim}{Y}_{n(i)}$. Hence, $m_{n(i)}$ is linearly dependent on $\underset{\sim}{Y}_{n(i)}$ and so the model is not identified. Notice as well that the Theorem requires that $\mathrm{E}(\underset{\sim}{X}_i \mid \underset{\sim}{Y}_{n(i)})$ is a nonlinear function of $\underset{\sim}{Y}_{n(i)}$; this is analogous to the condition for identification of some interaction effect in Manski (1993), Proposition 1 and Corollary. (Manski's results have to do with the identification of either an endogenous or contextual effect in the presence of individual effects, but does not allow for identification between these two effects, whereas our result gives conditions under which the two group effects can be distinguished.)

Why is there this difference between the binary choice and the linear-in-means frameworks? The answer is that the binary choice framework imposes a nonlinear relationship between the group characteristics and the group behaviors whereas the linear-in-means model (of course) does the opposite. Intuitively, suppose that one moves an individual from one neighborhood to another and observes the differences in his behavior. If the characteristics and behaviors of the neighborhoods always move in proportion as one moves across neighborhoods, then clearly one could not determine the respective roles of the characteristics as opposed to the behavior of the group in determining individual outcomes. This

can never happen in the logistic binary choice case given that the expected average choice must be bounded between $-1$ and 1. So, for example, as one moves across a sequence of arbitrarily richer communities, the percentage of high school graduates cannot always increase proportionately with income.

One can develop analogous identification conditions for alternative information assumptions in the linear-in-means model. For example, suppose that $\underset{\sim}{\bar{X}}_{n(i)}$, the sample average of the individual characteristics in neighborhood $n(i)$, is known to all members of the neighborhood. In this case,

$$m_{n(i)} = \frac{k + c'\underset{\sim}{\bar{X}}_{n(i)} + d'\underset{\sim}{Y}_{n(i)}}{1 - J}. \tag{58}$$

This equation makes clear that if the elements of $\underset{\sim}{\bar{X}}_{n(i)}$ lie in the linear space spanned by $\underset{\sim}{Y}_{n(i)}$, then the linear-in-means model will not be identified. Hence, we have the following corollary.

**Corollary 3. Necessary conditions for identification in the linear in means model when $\underset{\sim}{Y}_{n(i)}$ and $\underset{\sim}{\bar{X}}_{n(i)}$ are observable**

If $\underset{\sim}{Y}_{n(i)}$ and $\underset{\sim}{\bar{X}}_{n(i)}$ are observable, then a necessary condition for identification in the linear-in-means model is that the dimension of the linear space spanned by $(1, \underset{\sim}{X}_i, \underset{\sim}{Y}_{n(i)}, \underset{\sim}{\bar{X}}_{n(i)})$ is at least $r + s + 2$.

Operationally, this corollary means that for the full information case, one needs one individual variable whose neighborhood level average is not an element of the individual behavioral equation. This average can then be used to instrument $m^e_{n(i)}$.

*iii.* **instruments for unobservable expectations**

The identification condition for the linear-in-means model suggests a set of instruments which may be used when $m^e_{n(i)}$ is not observable, is measured with error, etc. Specifically, replacing $m^e_{n(i)}$ with the projection of $\bar{\omega}_{n(i)}$, the sample average of behaviors in neighborhood $n(i)$ onto $H(\underset{\sim}{Y}_{n(i)}, \mathrm{E}(\underset{\sim}{X}_i \mid \underset{\sim}{Y}_{n(i)}))$, where $H(a,b)$ denotes the Hilbert space generated by the elements of vectors $a$ and $b$, will not affect our identification results so long as $dim(H(\underset{\sim}{Y}_{n(i)}, \mathrm{E}(\underset{\sim}{X}_i \mid \underset{\sim}{Y}_{n(i)})) \ominus H(\underset{\sim}{Y}_{n(i)})) > 0$, where for Hilbert spaces $I$ and $G$ such that $G \subseteq I$, $I \ominus G$ denotes the Hilbert space generated by those elements of $I$ that are orthogonal to all elements of $G$. An analogous procedure will apply when $\bar{\underset{\sim}{X}}_{n(i)}$ is observable to individuals.

Of course, this assumes that the researcher has prior knowledge of what individual-level variables affect behavior when their neighborhood averages do not; otherwise, it would be the case that $H(\mathrm{E}(\underset{\sim}{X}_i \mid \underset{\sim}{Y}_{n(i)})) \subseteq H(\underset{\sim}{Y}_{n(i)})$ and so may be susceptible to Sim's (1980) classic critique of "incredible" identifying restrictions; see Freedman (1991) for a similar critique of the sorts of regressions we describe here. The point remains, however, that identification in the linear-in-means model depends on the same classical conditions as does identification in general simultaneous equations models, as initially recognized by Moffitt (1998).

At the same time, we would argue that the issue of omitted variables is far from insuperable. Both the social psychology and sociology literatures have focused a great deal of attention as to which types of individual and group control variables are most appropriate for inclusion in individual level regressions through the determination of which variables seem to be proximate versus ultimate causes of individual behavior; indeed it is this distinction which is the basis of path analysis (Blau and Duncan (1967)); see Sampson and Laub (1995) for what we consider a persuasive example of such a study. In general, we find it likely that these literatures will be able to identify examples of individual variables whose group average analogs are not proximate causes of behavior, and hence are available as instruments. While these literatures are often not driven by formal

statistical modelling and further subjected to Sims/Freedman-type critiques (e.g. Freedman (1991)) when formal techniques are employed, this hardly means that these literatures are incapable of providing useful insights. In this respect, we find arguments to the effect that because an empirical relationship has been established without justification for auxiliary assumptions such as linearity, exogeneity of certain variables, etc., one can ignore it, to be far overstated. In our view, empirical work establishes greater or lesser degrees of plausibility for different claims about the world and therefore the value of any study should not be reduced to a dichotomy between full acceptance or total rejection of its conclusions. Hence the determination of the plausibility of any exclusion restriction is a matter of degree and dependent on its specific context, including the extent to which it has been studied.

### *iv.* identification of individual versus neighborhood contextual effects

We now consider in more detail what is involved for identification of some type of neighborhood effects. What we mean is the following. Suppose that one wishes to determine whether any type of neighborhood effect exists, without distinguishing between endogenous and contextual effects, hence the only regressors in the model are a constant, $\underset{\sim}{X}_i$, and $\underset{\sim}{Y}_{n(i)}$. Operationally, we define this as determining whether, for a statistical model which only includes contextual effects as controls, the parameters on these contextual effects are identified.[2] A Corollary of the general identification Theorems 5 and 6 highlights the two conditions necessary to distinguish individual versus neighborhood contextual effects. A related result may be found in Manski (1993), corollary, pg. 535.

**Corollary 4. Identification of individual versus neighborhood effects in the binary**

[2]When $\underset{\sim}{Y}_{n(i)} = \mathrm{E}(\underset{\sim}{X}_i \mid i \in n(i))$ the corollary can be interpreted as applying to identification of a group effect for the reduced form of the linear-in-means model.

**choice model and linear-in-means model with global interactions**

In either the binary choice or the linear-in-means models with global interactions, a necessary condition for the identification of some neighborhood effect is that the dimension of the linear space spanned by the elements of $(1, \underset{\sim}{X}_i \underset{\sim}{Y}_{n(i)})$ is of higher dimension than the linear space spanned by the elements of $(1, \underset{\sim}{X}_i)$.

While the corollary is trivial, in that it is nothing more than the statement that in order to identify some sort of neighborhood effect some combination of the expected neighborhood effects must be linearly independent of the individual controls, it does have some economic content. Suppose that individuals are sorted into neighborhoods on the basis of an individual characteristic and that the neighborhood average of this same characteristic is what constitutes the relevant contextual effect. What this means is that there exists a neighborhood assignment rule $\xi(\cdot)$ which relates individual characteristics to neighborhood characteristics such that nonidentification requires that (assuming that the set of neighborhood characteristics and the set of individual characteristics are each internally linearly independent) there exists some linear combination of individual characteristics which is equal to some linear combination of neighborhood characteristics, *i.e.* there exist weights $\alpha$ and $\gamma$ such that

$$\kappa + \alpha' \underset{\sim}{X}_i = \gamma' \underset{\sim}{Y}_{n(i)} \tag{59}$$

But if this is so, then if individual observations are chosen randomly from the neighborhood, then it must be the case that individuals are perfectly segregated across neighborhoods with respect to the composite individual characteristic $\kappa + \alpha' \underset{\sim}{X}_i$. This is an extremely strong condition on the neighborhood sorting rule, ruling out any noise in the sorting process, and is in our judgment implausible. Hence our interpretation of the identification corollary is that empirical researchers should feel confident that individual versus neighborhood effects can

be at least in principle distinguished. To be clear however, this does not mean that data sets drawn from highly segregated communities are not a problem; rather the same reasoning we have applied suggests how segregation can lead to large standard errors for the estimated parameters of the model.

*v.* **nonlinear-in-means model**

The differences between the binary choice and linear-in-means models suggest that nonlinearity has a fundamental effect on the identification problem. McManus (1992) provides a number of general results which indicate that lack of identification is a nongeneric phenomenon in nonlinear contexts; these contexts do not include self-consistency conditions of the type which created the identification problem in the linear-in-means model. (A property is generic to a topological space of objects if it holds for an open, dense subset of the space.) McManus' analysis relies on some results from differential topology, which are beyond the scope of this chapter.

Nevertheless, it is possible to demonstrate a basic role for nonlinearity in identifying the parameters of interactions-based models by examining deviations from the linear-in-means model we have studied. Suppose that the individual behavioral equation is

$$\omega_i = k + c' \underset{\sim}{X}_i + d' \overline{\underset{\sim}{X}}_{n(i)} + J m^e_{n(i)} + \epsilon_i. \tag{60}$$

Here, the contextual effects $\overline{\underset{\sim}{X}}_{n(i)}$ are averages of the individual controls $\underset{\sim}{X}_i$, so we know that this model is not identified by Theorem 6. In the spirit of McManus (1992), we wish to make precise the idea that for the class of models, when the model is not linear in $m^e_{n(i)}$ but rather is linear in a function of $m^e_{n(i)}$, lack of identification is pathological.

To do this, let $g(m)$ be a $C^2$ function such that $g$ is nonlinear in $m$ and let

$$G(m^e_{n(i)}) = m^e_{n(i)} + \xi g(m^e_{n(i)}) \tag{61}$$

represent a class of functions which are perturbations around the linear function $m^e_{n(i)}$. We consider the nonlinear-in-means model

$$\omega_i = k + c' \underset{\sim}{X}_i + d' \bar{\underset{\sim}{X}}_{n(i)} + JG(m^e_{n(i)}) + \epsilon_i. \tag{62}$$

Associated with this equation is a conditional mean function $H$

$$H(\underset{\sim}{X}_i, \bar{\underset{\sim}{X}}_{n(i)}, m^e_{n(i)}) = k + c' \bar{\underset{\sim}{X}}_i + d' \bar{\underset{\sim}{X}}_{n(i)} + JG(m^e_{n(i)}). \tag{63}$$

For this model, self-consistency of $m^e_{n(i)}$ requires

$$m^e_{n(i)} = m_{n(i)} = k + (c' + d') \bar{\underset{\sim}{X}}_{n(i)} + JG(m_{n(i)}). \tag{64}$$

Our goal is to determine whether the model with $\xi = 0$ is special in terms of nonidentifiability of the parameters in (60). In doing so, we will assume that when there are multiple solutions to this equation, there is a selection rule which selects a particular solution $m_{n(i)}$ so that the observed $m_{n(i)} = m(\bar{\underset{\sim}{X}}_{n(i)})$.

In analyzing this equation, we will work with a notion of local identification. The model eqs. (61)-(64) define a "structure" for each particular parameter vector $A = (k,c,d,J)$. We focus here on identification at the level of the conditional mean function (63). Following Rothenberg (1971) or McManus (1992), we say a parameter point $A_0$ is locally identified if it fulfills the following definition. In our context, this condition is equivalent to requiring that the gradient vector of (63) with respect to $A$ has full rank.

**Definition. Local identification in the nonlinear-in-means model with interactions and self-consistent beliefs**

For the model described by eqs. (61)-(64), the parameter vector $A_0$ is locally identified if there exists an open neighborhood $N_{A_0}$ of $A_0$ such that no other parameter vector in $N_{A_0}$ gives the same conditional mean in eq. (63) and such that the self-consistency condition eq. (64) holds as well.

The concept of local identifiability has value as argued in Rothenberg (1971) pg. 578 as

"It is natural to consider the concept of local identification. This occurs when there may be a number of observationally equivalent structures but they are isolated from each another."

Rothenberg (1971) demonstrates that there is a close connection between local identification and the full rank assumption of particular derivative matrices of a likelihood function. In our context, this means that one must show that the gradient of the conditional mean function with respect to $A$ is of full rank. In addition, we need to account for the self-consistency condition in the sense that the full rank condition must hold when the gradient is evaluated at a solution $m_{(i)}$ to the self-consistency condition. The following Theorem is verified in the Mathematical Appendix.

**Theorem 7. Local identifiability for models in a neighborhood of the linear-in-means model**

Assume

*i.* $supp(\underset{\sim}{\bar{X}}_{n(i)})$ is not contained in a proper linear subspace of $R^r$.

*ii.* There exists at least one neighborhood $n_0$ such that conditional on $\underset{\sim}{\bar{X}}_{n_0}$, $\underset{\sim}{X}_i$ is not contained in a proper linear subspace of $R^r$.

*iii.* $J \neq 1$.

*iv.* The population data $\{\bar{\underset{\sim}{X}}_{n(i)}, \underset{\sim}{X}_i, m_{n(i)}\}$ is such that there is an open set $O$ such that $m(\bar{\underset{\sim}{X}}_{n(i)})$ is differentiable on $O$ and nonconstant on $O$. Further, there are two distinct values in $O$, call them $\bar{\underset{\sim}{X}}_1$ and $\bar{\underset{\sim}{X}}_2$, such that $m_1 = m(\bar{\underset{\sim}{X}}_1) \neq m_2 = m(\bar{\underset{\sim}{X}}_2)$ and $\frac{dg(m_1)}{dm} \neq \frac{dg(m_2)}{dm}$.

Then there exists an open neighborhood $N$ of $\xi = 0$, such that $\forall\ \xi \in\ N - \{0\}$, the model defined by eqs. $(61) - (64)$ is locally identified.

What is important about this theorem is that it highlights the importance of linearity in generating nonidentification. For a permutation of the linear in means model in the direction of any nonlinear function $g$, identification will hold. As nonlinearity seems to be a very standard feature of models with interactions, this result provides a relatively optimistic perspective on the identification problem, at least for the case of correctly specified models.

We believe that it should be relatively straightforward to extend the approach of McManus to show that identification is a generic property of nonlinear models with self-consistency constraints and are pursuing this in subsequent work.

***vi.* implications of self-selection for identification**

Our discussion thus far has assumed that the rules by which individuals are sorted into groups has no implications for empirical analysis. Such an assumption implies that the group formation rule is independent of the determinants of individual choices and is thus unnatural in many contexts. Given the preferences we have assumed, one would expect individuals, when possible, to endogenously

sort themselves, accounting for the effects of neighborhood characteristics and expected neighborhood behavior on payoff functions. Hence, there is the potential for self-selection bias. For decisions such as nonmarital births or dropping out of school, standard estimation methods may produce biased estimates due to the correlation of the $\epsilon_i(\omega_i)$'s with the determinants of sorting. To be clear, we do not explicitly account for equilibrium group formation, but rather approximate its effects through consideration of selection.

This issue has yet to be addressed in an extended fashion in the interactions literature. With reference to identification, what appears important is that self-selection may actually facilitate identification. Intuitively, self-selection can induce precisely the sort of nonlinearities or exclusion restrictions which generates identification in the earlier discussion.

To see this, we develop an example. Suppose that the econometric version of the linear-in-means model, eq. (55), describes the behavioral rule for all individuals in a population, but that we only observe those outcomes for individuals who have been sorted into neighborhoods in the sample. This can be justified by positing the existence of a reservation neighborhood for each individual. We assume that this means that there is a latent variable $z_i$ which measures a family's evaluation of the neighborhood and such that a family is observed in neighborhood $n(i)$ if and only if $z_i > 0$. In turn, this latent variable can be written as

$$z_i = \gamma' \underset{\sim}{R}_i + \eta_i \tag{65}$$

where $\underset{\sim}{R}_i$ is a vector of determinants of $i$'s neighborhood evaluation. Finally, assume that the errors $\epsilon_i$ and $\eta_i$ are zero mean, jointly normal with the variance/covariance matrix

$$\begin{bmatrix} \sigma_\epsilon^2 & \rho\sigma_\epsilon \\ \rho\sigma_\epsilon & 1 \end{bmatrix} \tag{66}$$

and where $\mathrm{E}(\epsilon_i \mid \underset{\sim}{X}_i, \underset{\sim}{Y}_{n(i)}, m^e_{n(i)}, \underset{\sim}{R}_i) = \mathrm{E}(\eta_i \mid \underset{\sim}{X}_i, \underset{\sim}{Y}_{n(i)}, m^e_{n(i)}, \underset{\sim}{R}_i) = 0$.

This is precisely the model which is considered in Heckman (1979). Following his argument, since

$$\mathrm{E}(\epsilon_i \mid z_i > 0) = \rho\sigma_\epsilon \lambda_i(\gamma' \underset{\sim}{R}_i) \tag{67}$$

where, letting $\phi(\cdot)$ and $\Phi(\cdot)$ respectively denote the standard normal density and distribution,

$$\lambda(\gamma' \underset{\sim}{R}_i) = \frac{\phi(\gamma' \underset{\sim}{R}_i)}{\Phi(\gamma' \underset{\sim}{R}_i)}, \tag{68}$$

a regression in which the model disturbance is orthogonal to the various regressors is

$$\omega_i = k + c' \underset{\sim}{X}_i + d' \underset{\sim}{Y}_{n(i)} + J m^e_{n(i)} + \rho\sigma_\epsilon \lambda(\gamma' \underset{\sim}{R}_i) + \zeta_i. \tag{69}$$

What is important for our purposes is that the structure of this equation can facilitate identification. There are two distinct ways in which this can occur.

First, consider the case where each individual control is matched one-to-one with a contextual effect so that $\mathrm{E}(\underset{\sim}{X}_i \mid \underset{\sim}{Y}_{n(i)}) = \underset{\sim}{Y}_{n(i)}$. Assume as well that none of the variables in $\underset{\sim}{R}_i$ are functionally dependent on $m^e_{n(i)}$, so that we may assume that the reduced form for $m^e_{n(i)}$ depends on $\underset{\sim}{R}_i$. As discussed above, if $\rho\sigma_\epsilon = 0$, so there is no selection correction, this is Manski's (1993) nonidentification example. However, in the presence of self-selection, the expected average choice within a neighborhood is, under self-consistency

$$m_{n(i)} = \frac{k}{1-J} + \frac{1}{1-J}(c' + d')\underset{\sim}{Y}_{n(i)} + \frac{\rho\sigma_\epsilon}{1-J}\mathrm{E}(\lambda(\gamma'\underset{\sim}{R}_i) \mid i \in n(i)) \tag{70}$$

so that a reduced form for individual behavior may be written as

$$\omega_i =$$

$$\frac{k}{1-J} + c'\underset{\sim}{X}_i + \frac{1}{1-J}(Jc' + d')\underset{\sim}{Y}_{n(i)} +$$

$$\rho\sigma_\epsilon\lambda(\gamma'\underset{\sim}{R}_i) + \frac{J\rho\sigma_\epsilon}{1-J}\mathrm{E}(\lambda(\gamma'\underset{\sim}{R}_i) \mid i \in n(i)) + \zeta_i. \tag{71}$$

In this reduced form regression, nonidentification when $\rho\sigma_\epsilon = 0$ follows immediately from observing that there are $2r+1$ parameters and only $2r$ regressors. However, when there is a selection correction, two new regressors are introduced, $\lambda(\gamma'\underset{\sim}{R}_i)$ and $\mathrm{E}(\lambda(\gamma'\underset{\sim}{R}_i) \mid i \in n_i)$, but only one new parameter, $\rho\sigma_\epsilon$. This allows for identification so long as $\lambda(\gamma'\underset{\sim}{R}_i)$ and $\mathrm{E}(\lambda(\gamma'\underset{\sim}{R}_i) \mid i \in n_i)$ are not perfectly collinear, which requires that there is within-neighborhood variation in $\lambda(\gamma'\underset{\sim}{R}_i)$. Notice that the nonlinearity of $\lambda(\,\cdot\,)$ ensures that the appearance of regressors in $\underset{\sim}{R}_i$ which appear in either $\underset{\sim}{X}_i$ or $\underset{\sim}{Y}_{n(i)}$ does not imply nonidentification due to multicollinearity of the correction term with the other variables in the model.

This route to identification through selection correction is an example of the general identification condition stated in Theorem 6. In order to achieve identification, one needs an individual control whose neighborhood average is not a contextual effect. This is precisely what occurs when $\lambda(\gamma'\underset{\sim}{R}_i)$ is introduced into the linear-in-means model, since $\mathrm{E}(\lambda(\gamma'\underset{\sim}{R}_i) \mid i \in n_i)$ is not an element of the model even when selection is controlled for.

Second, identification may be achieved if $m_{n(i)}$ is a component of $\underset{\sim}{R}_i$. Suppose that the expected average choice level is the only element in $\underset{\sim}{R}_i$. The selection-corrected linear-in-means model is now

$$\omega_i = k + c'\underset{\sim}{X}_i + d'\underset{\sim}{Y}_{n(i)} + Jm^e_{n(i)} + \rho\sigma_\epsilon\lambda(\gamma m_{n(i)}) + \zeta_i. \tag{72}$$

The parameters in this regression will now be identified so long as the joint support of $\underset{\sim}{X}_i$ and $\underset{\sim}{Y}_{n(i)}$ does not lie in a proper linear subspace of $R^{r+s}$ since the nonlinearity of the selection correction ensures that there is no linear dependence between $m^e_{n(i)}$ and the individual and neighborhood controls. Notice that this is the same reason for identification derived for the binary choice model; in both cases, the nonlinear dependence of $m^e_{n(i)}$ on $\underset{\sim}{X}_i$ and $\underset{\sim}{Y}_{n(i)}$ produces identification.

Of course, identifiability of model parameters does not say anything about the precision of the estimates facilitated by selection corrections. Intuitively, one will need substantial cross-neighborhood variation in $m^e_{n(i)}$ if the nonlinear dependence of the correction on this term is the basis for identification. Similarly, substantial variation in $\underset{\sim}{R}_i$ will be needed if elements of this vector are highly correlated with combinations of $\underset{\sim}{X}_i$ and $\underset{\sim}{Y}_{n(i)}$, in order for the nonlinearity of the correction to avoid multicollinearity. Notice in this case, the presence of regressors in $\underset{\sim}{R}_i$ which do not appear in $\underset{\sim}{X}_i$ or $\underset{\sim}{Y}_{n(i)}$ will likely prove valuable in practice.

To be clear, this discussion hardly exhausts the implications of selection corrections for identification. One issue concerns the relationship between the selection and behavior equations. There is no behavioral justification for the selection equation we have employed whereas ideally the selection equation will reflect individual optimization over a set of neighborhood choices and account for subsequent behavior which will occur in the neighborhood. Further, the analysis needs to be extended to cases where the joint normality of the selection and behavior disturbances is relaxed. Examples of nonparametric approaches to selection correction include Ahn and Powell (1993). What this example nevertheless demonstrates is that self-selection can, when accounted for, work to aid in identification, and hence clearly warrants further research.

*vii.* **implications of multiple equilibria for identification**

Finally, we observe that contrary to much of the conventional wisdom, the presence of multiple steady states can provide identification in and of itself, a possibility suggested in Manski (1993) pg. 539. To see this, suppose that all neighborhoods are composed of individuals with identical characteristics, so that $y_{n(i)} = \overline{y}$. Suppose that the $J$ is greater than 1 and that $d\overline{y}$ is small enough relative to $J$ that there are multiple steady states in a neighborhood. Finally, suppose that a fraction $r$ of neighborhoods exhibit expected average choices consistent with the largest solution to $m = tanh(d\overline{y} + Jm)$, and a fraction $1-r$ exhibit average choices consistent with the smallest solution. In this case the determinant of the covariance matrix of $y_{n(i)}$ and $m_{n(i)}$ is $\overline{y}^2 var(m_{n(i)})$ which is nonzero unless $\overline{y}$ is zero. This would imply that in a regression of the form

$$\omega_i = k + d\underset{\sim}{Y}_{n(i)} + Jm^e_{n(i)} + \epsilon_i \tag{73}$$

$J$ will be identified (although $k$ and $d$ of course will not be). The intuitive point is that variation in the realized equilibria across observations for a model with multiple equilibria can provide the leverage required to identify model parameters.

### *viii.* dynamic models and rational expectations

Wallis (1980) provides an analysis of identification in classical rational expectations models that is closely related to the analysis of identification in the linear-in-means model. Suppose that the linear-in-means model is modified so that it now describes behavior at points in time, i.e.

$$\omega_{i,t} = c'\underset{\sim}{X}_{i,t} + d'\underset{\sim}{Y}_{n(i),t} + Jm^e_{n(i),t} + \epsilon_{i,t}. \tag{74}$$

Let $\underset{\sim}{\omega}_t$ denote the column vector of choices at $t$, $X_t$ and $Y_t$ denote matrices whose columns are the $\underset{\sim}{X}_{i,t}$'s and $\underset{\sim}{Y}_{n(i),t}$'s respectively, and $C$ and $D$ denote conformable matrices whose rows are always $c'$ and $d'$ respectively. Then a panel of observations on individuals can be written as

$$\underset{\sim}{\omega}_t = CX_t + DY_t + J\underset{\sim}{m}^e_t + \underset{\sim}{\varepsilon}_t. \tag{75}$$

When $D = 0$ and $J$ is not a scalar, but rather a conformable matrix, one has the vector linear – in – means model version of the Wallis (1980) structural equation (2.1). These differences between the linear-in-means model and Wallis' model create new problems for identification. The identification problems which we have described in the linear-in-means model occur precisely because of the need to identify $D$. The identification problem is particularly acute when the $Y$ matrix consists of neighborhood averages of $X_t$, as we have already seen.

Observing the connection between the linear-in-means model and (75) suggests that fruitful connections exist between the literature we survey here and the classical rational expectations econometric work of Hansen and Sargent (1991) and Wallis (1980), among many others, as well as more recent work that extends that tradition to social interactions (Binder and Pesaran (1997)) and to spatial rational expectations econometrics (Fingleton (1999)).

As eqs. (74) and (75) make clear, the linear-in-means model is interpretable as a version of the Wallis model where $\underset{\sim}{\omega}_t$ is scalar. The identification problem which occurs when $m_{n(i)}$ can be expressed as a linear combination of $\underset{\sim}{Y}_{n(i)}$'s will occur in Wallis' model when $\underset{\sim}{Y}_{n(i)} = \mathrm{E}(\underset{\sim}{X}_i \,|\, \underset{\sim}{Y}_{n(i)})$ for the various columns of $Y_t$. This connection between the problem of identification in the linear-in-means model which describes interactions in "space" with the problem of identification in linear rational expectations models in "time" suggests integrative future research along these lines should exist.

There are in fact many dimensions along which one can explore links between interactions-based models and rational expectations models. Hansen and

Sargent (1991), pg. 2 remark that

"Work on rational expectations econometrics has divided into two complementary but differing lines. The first line aims more or less completely to characterize the restrictions that a model imposes on a vector stochastic process of observables, and to use those restrictions to guide efficient estimation. This line is a direct descendant of the full system approach to estimating simultaneous equation models...

The second line of work is the application of method of moments estimators to estimating the parameters that appear in the Euler equations associated with dynamic optimization problems..."

Our discussion thus far has contrasted the linear-in-means model of social interactions with what Hansen and Sargent call the "first line" which is treated by Wallis in a framework particularly suited to comparison. In spatial optimization problems one could also develop a "second" line that parallels the Euler equation-based, methods of moments approach.

A key feature of dynamic rational expectations models is the potential for intertemporal interaction effects to influence identifiability. For example, suppose that individuals are affected by lagged group characteristics and lagged expected average behavior, so that

$$\omega_{i,t} = c' \underset{\sim}{X}_{i,t} + d' \underset{\sim}{Y}_{n(i),t-1} + J m^{e}_{n(i),t-1} + \epsilon_{i,t}. \tag{76}$$

(We omit the constant $k$ for expositional reasons.) In the case where all individual characteristics correspond one to one with neighborhood contextual effects (which is the Manski case of no identification in the linear-in-means model), this equation can be reexpressed as

$$m_{n(i),t} = c' \underset{\sim}{Y}_{n(i),t} + d' \underset{\sim}{Y}_{n(i),t-1} + J m_{n(i),t-1} \tag{77}$$

or

$$m_{n(i),t} = (\frac{1}{1-JL})c'\underset{\sim}{Y}_{n(i),t} + (\frac{1}{1-JL})d'\underset{\sim}{Y}_{n(i),t-1}. \tag{78}$$

where $L$ denotes a lag operator. (We have assumed $|J| < 1$ so that the operator $1 - JL$ is invertible.) Substituting this expression into (76),

$$\omega_{i,t} =$$

$$c'\underset{\sim}{X}_{i,t} + d'\underset{\sim}{Y}_{n(i),t-1} + J[(\frac{1}{1-JL})c'\underset{\sim}{Y}_{n(i),t-1} + (\frac{1}{1-JL})d'\underset{\sim}{Y}_{n(i),t-2})] + \epsilon_{i,t} =$$

$$c'\underset{\sim}{X}_{i,t} + (Jc' + d')\underset{\sim}{Y}_{n(i),t-1} + (\frac{J}{1-JL})(Jc' + d')\underset{\sim}{Y}_{n(i),t-2} + \epsilon_{i,t} \tag{79}$$

where the last line in the equation follows from $(\frac{1}{1-JL})x_t = x_t + (\frac{J}{1-JL})x_{t-1}$. Now, assume that the moment matrix generated by the elements of $(\underset{\sim}{X}_{i,t}, \underset{\sim}{Y}_{n(i),t-1}, \underset{\sim}{Y}_{n(i),t-2}, \ldots)$ has full rank, so that the coefficient on each of the variables on the right hand side of (79) is identified. Then the coefficients of the underlying structural model are also identified. To see this, observe first that $c$ is of identified by the coefficients on the regressors $\underset{\sim}{X}_{i,t}$. $J$ is identified because the coefficients on any corresponding elements of $\underset{\sim}{Y}_{n(i),t-k}$ and $\underset{\sim}{Y}_{n(i),t-k-1}$ with $k > 1$ proportional to $J$. Once $c$ and $J$ are identified, so is $d$ from the coefficients on any set of regressors $\underset{\sim}{Y}_{n(i),t-1}$. Intuitively, the timing of the interactions breaks the strict collinearity of the contextual and endogenous effects.[3]

Finally, as an example of how the substantive economics in a dynamic model can influence identification, we consider a dynamic model of production complementarities of the type studied by Binder and Pesaran (1997). In this model, the capital decisions of a set of profit maximizing firms is studied. Each firm possesses a technology such that

[3]Manski (1993) pg. 540 conjectures that a lagged linear-in-means model may be identified. Our verification of this conjecture suggests that the reason is not that the data are out of "temporal equilibrium" as Manski suggests, but rather that the collinearity of expected group outcomes and contextual effects is affected by dynamics in the interactions.

$$Q_{i,t} = T_{i,t} K_{i,t}^{\alpha} \tag{80}$$

where $T_{i,t}$ measures the level of firm $i$'s technology and $K_{i,t}$ measures its capital stock. The rental price of capital for each firm is $R_t$. The level of technology of firm $i$ is assumed to follow

$$T_{i,t} =$$

$$Aexp(c' log(\underset{\sim}{X}_{i,t}) + d' \mathrm{E}(log \underset{\sim}{X}_{i,t} \mid i \in n(i)) + J \mathrm{E}(log K_{i,t} \mid i \in n(i)) + \epsilon_{i,t}) \tag{81}$$

In this formulation, for any $\underset{\sim}{w}$, $log \underset{\sim}{w}$ is the vector whose $i$'th element is $log w_i$. The shock $\epsilon_{i,t}$ is taken to be independent and identically distributed across both firms and time.

Firms are assumed to maximize the expectation of the present discounted value of their current and future profits. Each firm observes its own shock $\epsilon_{i,t}$ at the time it chooses $K_{i,t}$ but does not observe the shocks of other firms. Given our assumption that the technology shocks are independent across time, the discounted sum of profits breaks down into a sum of independent profit terms. Profit maximization with respect to the choice of firm-specific capital leads to the first-order condition

$$(1-\alpha) log K_{i,t} =$$

$$log(\alpha A) + c' log(\underset{\sim}{X}_{i,t}) + d' \mathrm{E}(log \underset{\sim}{X}_{i,t} \mid i \in n(i)) + J \mathrm{E}(log K_{i,t} \mid i \in n(i))$$

$$- log R_t + \epsilon_{i,t} \tag{82}$$

Suppose that we have data for a cross-section of firms at fixed $t$. In this case, (82) is an example of the linear-in-means model for which identification fails, since the group analog of each individual control appears in the structural

equation; specifically, we have $r$ individual controls $log(\underset{\sim}{X}_{i,t})$ and $r$ group level controls $\mathrm{E}(log\underset{\sim}{X}_{i,t} \mid i \in n(i))$, and the composite constant term $log(\alpha) - logR_t$, so by Theorem 6, the model is not identified.[4]

Alternative routes to identification emerge when one allows for a richer dynamic structure to technology. Productivity spillovers generated by one firm onto another plausibly depend only on the current level of that firm's technology, not the particular path by which the firm arrived at that technology. On the other hand, the ability of any firm $i$ to benefit from another firm's technology plausibly will depend on its own level of technology in the previous period as well as its characteristics today.

Formally, these assumptions on the dynamics of spillovers can be expressed in an equation for the logarithm of firm $i$'s technology level such as

$$T_{i,t} =$$

$$T_{i,t-1}^{\beta} A exp(\mathrm{c}' log(\underset{\sim}{X}_{i,t}) + d' \mathrm{E}(log\underset{\sim}{X}_{i,t} | i \in n(i)) + J\mathrm{E}(logK_{i,t} | i \in n(i)) + \epsilon_{i,t}) \quad (83)$$

The first order conditions for profit maximization now imply, after taking $log$'s, that the capital level for each firm obeys

$$(1-\alpha) logK_{i,t} =$$

$$\beta logT_{i,t-1} + log(\alpha A) + c' log(\underset{\sim}{X}_{i,t}) +$$

$$d' \mathrm{E}(log\underset{\sim}{X}_{i,t} \mid i \in n(i)) + J\mathrm{E}(logK_{i,t} \mid i \in n(i)) - logR_t + \epsilon_{i,t}. \quad (84)$$

What matters from the perspective of identification is that we now have an additional regressor $logT_{i,t-1}$, whose group average does not appear in the

[4]Notice that we do not assume in this particular case that the averages of the individual characteristics $log\underset{\sim}{X}_{i,t}$ are known by members of a neighborhood. This has no effect on the analysis.

equation. Hence for this model, we have $r+1$ individual effects, whereas we still have only $r$ contextual effects. Hence, it will be possible to identify $c$, $d$, $J$. Of course, if this variable is not observable, it will itself have to be instrumented.

This example is only meant to be illustrative. Once one leaves the log linear framework, recent work such as Pakes (1999) that treats nonlinearities seriously in firm dynamics would be needed for the formulation of the stochastic processes characterizing firm behavior. Extending this kind of analysis to focus on measuring spillovers between firms seems to us to be a worthwhile area for future research. While production spillovers are the driving force behind the new growth theory, it is remarkable how little firm evidence of such spillovers actually exists. A primary reason for this is the weakness of the econometric methods which have been employed to obtain such evidence; Durlauf and Quah (1999) discuss many of the problems which exist with efforts to identify production externalities using cross-country growth data. Our belief is that the use of individual level data which follows interactions in the way we have described will yield much clearer inferences.

## 4. Further topics in identification

### *i.* panel data

The extension of identification results from cross-sections to panels is important for several reasons. First, panels will provide an opportunity for dealing with model misspecification which is not present in cross-sections. Hoffman and Plotnick (1996) is the only case we are aware of in which this argument is applied in an interactions context. Second, panels allow for intertemporal interactions which facilitate a richer notion of belief formation than we have used.

*a.* fixed effects

To see how panel data can provide a way of dealing with misspecification, we start with the linear-in-means case. The panel analog to eq. (55) is

$$\omega_{i,t} = k + c'\underset{\sim}{X}_{i,t} + d'\underset{\sim}{Y}_{n(i,t),t} + Jm^{e}_{n(i,t),t} + \alpha_i + \epsilon_{i,t}. \tag{85}$$

In addition to introducing time subscripts on all variables, an unobservable fixed effect term $\alpha_i$ is now included.

In order to produce consistent estimates of $c$, $d$, and $J$, we follow the suggestion of Chamberlain (1984) and difference this equation with respect to $t$,

$$\Delta\omega_{i,t} = c'\Delta\underset{\sim}{X}_{i,t} + d'\Delta\underset{\sim}{Y}_{n(i,t),t} + J\Delta m^{e}_{n(i,t),t} + \Delta\epsilon_{i,t}. \tag{86}$$

Identification may now be treated in a fashion exactly analogous to that in Section 3, once first differences replace the levels used in the identification conditions. Notice that it is important to be careful about the assumption that regressors are orthogonal to errors in the differenced equation

$$\mathrm{E}(\Delta\epsilon_{i,t} \mid \Delta\underset{\sim}{X}_{i,t}, \Delta\underset{\sim}{Y}_{n(i,t),t}, \Delta m^{e}_{n(i,t),t}) = 0$$

since $\Delta\epsilon_{i,t}$ will not be white noise.

One implication of the panel data case with fixed effects is that variation in $\Delta m^{e}_{n(i,t),t}$ is useful in facilitating identification. In turn, this suggests that in environments which are slowly moving, $J$ may be difficult to estimate precisely.

Analogous reasoning may be applied to the binary choice case. Suppose that the individual payoff function is now

$$V(\omega_{i,t},\underset{\sim}{Z}_{i,t},\mu^e_{i,t}(\underset{\sim}{\omega}_{-i,t}),\epsilon_{i,t}(\omega_{i,t}),\alpha_i) \tag{87}$$

where we again introduce time subscripts and a fixed effect $\alpha_i$. We generalize our earlier development of the individual choice problem by assuming that the differential payoff between the two choices equals

$$c'\underset{\sim}{X}_{i,t} + d'\underset{\sim}{Y}_{n(i,t),t} + Jm^e_{n(i,t),t} + \epsilon_{i,t}(1) - \epsilon_{i,t}(-1) + \alpha_i \tag{88}$$

So that the probability measure for $\omega_{i,t}$ obeys

$$\mu(\omega_{i,t} = 1 \mid \underset{\sim}{X}_{i,t},\underset{\sim}{Y}_{n(i,t)},m^e_{n(i,t)}) \sim$$

$$exp(\beta c'\underset{\sim}{X}_{i,t} + \beta d'\underset{\sim}{Y}_{n(i,t)} + \beta Jm^e_{n(i,t),t} + \alpha_i) \tag{89}$$

This equation is in a form which is estimable given methods derived in Honoré and Kyriazidou (1998). That paper also shows how one can adapt Manski's (1975,1985) maximum score estimator so as to allow estimation of the model's parameters without assuming a logistic distribution for differences in the random utility terms. With respect to the logistic case, Honoré and Kyriazidou (1998) show, extending an original insight of Chamberlain (1993), that if one has two consecutive observations on agent $i$, and if lagged $\omega_{i,t}$'s do not appear in the regressor matrix, then the conditional probability that either $\omega_{i,t-1} = -1$ or $\omega_{i,t} = 1$ given $\omega_{i,t-1} + \omega_{i,t} = 0$ is independent of $\alpha_i$. This is the analogy to the differencing out of the fixed effect in the linear-in-means case. Honoré and Kyriazidou further show that if lagged $\omega_{i,t}$'s do appear in the regressor matrix, it is possible to modify their procedures and achieve identification so long as four consecutive observations on each agent are available. With respect to their conditions for identification, they appear to be consistent with the interactions-models we have been analyzing, once one allows for differenced rather than levels

data.

This discussion has of course assumed that there is no self-selection into groups. Kyriazidou (1997a,b) provides conditions under which identification can occur for this case; extension of her methods to interactions-based environments would seem quite valuable.

### *b.* learning

An alternative use of panel data lies in the ability to model the expectations process as generated by learning. A simple way of doing this is to assume that the beliefs of an individual concerning the average choice in the population equals the realized average last period, so that social utility, for example, may be written as

$$S(\omega_{i,t}, \underset{\sim}{X}_{i,t}, \underset{\sim}{\omega}_{-i,t-1}) = -\sum_{j \neq i} \frac{J_{i,j}}{2} (\omega_{i,t} - \omega_{j,t-1})^2 \tag{90}$$

At this level, the dynamic interactions can be interpreted either as reflecting a primitive assumption either about individual preferences or concerning the way in which individuals form expectations of the contemporaneous behavior of others. This will not be so for more sophisticated learning models with interaction effects. Such models have been studied by Besley and Case (1994), Case (1995) and Munshi and Myaux (1998). In complementary work, Binder and Pesaran (1998a) provide an interesting analysis in which social interactions can be analyzed in a dynamic model with rational expectations.

## *ii.* duration data

For contexts such as out-of-wedlock births or first sexual activity, it seems natural to consider interactions as they affect the probability of transition from one state to another; see Brewster (1994a,b) and Sucoff and Upchurch (1998) for empirical studies using this perspective. For such models, it is necessary to reformulate the nature of the interactions as they are manifested in a self-consistency condition analogous to eq. (16) in order to exploit the tools which have been developed to study duration data; these tools are well surveyed in Cox and Oakes (1984), Heckman and Singer (1984a, 1985) and Lancaster (1990). To keep matters concrete, we will use the timing of out-of-wedlock births as an example.

Following standard notation (e.g. Amemiya (1985) Section 11.2 or Heckman and Singer (1985a)), we let $T$ denote the duration from $t = 0$ that an unmarried woman remains childless. The probability that this duration is less than any $t$, $\mu(T < t)$, is denoted as $F(t)$. For any interval $\delta t$, the probability that a childless woman at $t$ becomes pregnant by $t + \delta t$ is

$$\mu(t \leq T < t + \delta t \mid T \geq t) = \frac{\mu(t \leq T < t + \delta t, T \geq t)}{\mu(T \geq t)}. \tag{91}$$

From this condition probability, two standard functions of interest can be defined. First, the hazard function $\lambda(t)$ is defined as

$$\lambda(t) = lim_{\delta t \Rightarrow 0} \frac{\mu(t \leq T < t + \delta t, T \geq t)}{\delta t \mu(T \geq t)} = \frac{F'(t)}{1 - F(t)}. \tag{92}$$

Second, the survivor function $S(t)$ is defined as $1 - F(t)$. If $\lambda(t) = \lambda$, so that the hazard is independent of time, then the survivor function is

$$S(t) = exp(-\lambda t), \tag{93}$$

which is the standard exponential form employed in many applications.

In order to make clear the basic identification issues we start with a

baseline case. We assume that the time scale of the duration of interest is short relative to the time scale over which data are collected, so that all duration "spells" are completed. Formally, this assumption means that there is no "right censoring" of the data. This will not be appropriate for out-of-wedlock birth data, since of course not all unmarried females experience the event; nevertheless, the assumption is useful for exposition. Let $t_i$ denote the time of first birth for individual $i$. If this timing is associated with probability density $f(\cdot)$, then the joint density for the $I$ times is $\prod_i f(t_i)$. This joint probability will be determined by the individual hazards $\lambda_i$.

As before, we assume that the hazard for each individual under analysis depends on individual characteristics $\underset{\sim}{X}_i$, neighborhood characteristics $\underset{\sim}{Y}_{n(i)}$ and an expected neighborhood behavioral measure $m^e_{n(i)}$. In this context, $m^e_{n(i)}$ may be the expected value of either the within-neighborhood duration or the median group duration. We therefore assume that for each individual

$$\lambda_i = \lambda(\underset{\sim}{X}_i, \underset{\sim}{Y}_{n(i)}, m^e_{n(i)}) \tag{94}$$

so that the associated density for the duration is

$$f(t \mid \underset{\sim}{X}_i, \underset{\sim}{Y}_{n(i)}, m^e_{n(i)}) = \lambda(\underset{\sim}{X}_i, \underset{\sim}{Y}_{n(i)}, m^e_{n(i)}) exp(-\lambda(\underset{\sim}{X}_i, \underset{\sim}{Y}_{n(i)}, m^e_{n(i)})t) \tag{95}$$

The expected duration for individual $i$, conditional on these controls is

$$\mathrm{E}(t \mid \underset{\sim}{X}_i, \underset{\sim}{Y}_{n(i)}, m^e_{n(i)}) = \lambda(\underset{\sim}{X}_i, \underset{\sim}{Y}_{n(i)}, m^e_{n(i)})^{-1} \tag{96}$$

and the median of the duration is given by the solution $t^*$ to $F(t^*) = \frac{1}{2}$ which implies that

$$\tfrac{1}{2} = 1 - F(t^*) = exp(-\lambda(\underset{\sim}{X}_i, \underset{\sim}{Y}_{n(i)}, m^e_{n(i)})t^*) \tag{97}$$

which implies that $t^*$ solves

$$log2 = \lambda(\underset{\sim}{X}_i, \underset{\sim}{Y}_{n(i)}, m^e_{n(i)})t^* \tag{98}$$

Equations (97) and (98) allow us to define self-consistent solutions for this model. Self-consistency with respect to expected duration times requires that

$$m^e_{n(i)} = m_{n(i)} = \int \lambda(\underset{\sim}{X}_i, \underset{\sim}{Y}_{n(i)}, m_{n(i)})^{-1} dF_{\underset{\sim}{X}} \tag{99}$$

where as before $F_{\underset{\sim}{X}}$ is the probability distribution of characteristics within neighborhood $n(i)$. Similarly, self-consistency with respect to the neighborhood median requires that

$$m^e_{n(i)} = m_{n(i)} = log2 \int \lambda(\underset{\sim}{X}_i, \underset{\sim}{Y}_{n(i)}, m_{n(i)})^{-1} dF_{\underset{\sim}{X}}\,. \tag{100}$$

These two expressions only differ by a constant of proportionality. Notice that we have assumed that each individual references on her entire neighborhood. It is possible to consider cases where the reference group is smaller, so that for example, one only references on individuals with similar individual characteristics. As we have already seen in the discussion of other models, the "width" of each individual's reference group plays a key role in identification.

We first consider identification in the parametric case under the assumption that the expected value of the duration time within a group is the relevant endogenous interaction. Following treatments such as Amemiya (1985) section 11.2.3, we assume that the hazard function for individual $i$ is exponential, so that

$$\lambda_i = exp(c' \underset{\sim}{X}_i + d' \underset{\sim}{Y}_{n(i)} + J m^e_{n(i)}). \tag{101}$$

We assume that $\underset{\sim}{X}_i$ contains a constant term (Amemiya (1985) eq. 11.2.26). The

associated likelihood function for the data will therefore be

$$L = \prod_i exp(c'\underset{\sim}{X}_i + d'\underset{\sim}{Y}_{n(i)} + Jm^e_{n(i)})exp(-exp(c'\underset{\sim}{X}_i + d'\underset{\sim}{Y}_{n(i)} + Jm^e_{n(i)})t_i). \quad (102)$$

For this model, choosing parameter estimates for $c$, $d$, and $J$ to maximize (102) without imposing eq. (99) corresponds to the naive estimator we have described in the binary choice and linear-in-means cases.

Following the analysis in Amemiya (1985), identification in population requires that the expected value of the Hessian matrix of $logL$ is nonsingular at the self-consistent solution (99). Letting $b = (c',d',J)'$ and $\underset{\sim}{M}_i = (\underset{\sim}{X}'_i, \underset{\sim}{Y}'_{n(i)}, m^e_{n(i)})'$, the expected value of the Hessian equals

$$\mathrm{E}(\frac{\partial^2 logL}{\partial b \partial b'} \mid \underset{\sim}{M}_i) = -\mathrm{E}(\sum_i t_i exp(b'\underset{\sim}{M}_i)\underset{\sim}{M}_i\underset{\sim}{M}'_i \mid \underset{\sim}{M}_i). \quad (103)$$

Further, since $\mathrm{E}(t_i \mid \underset{\sim}{M}_i) = \lambda_i^{-1}$, it is further the case that

$$-\mathrm{E}(\sum_i t_i exp(b'\underset{\sim}{M}_i)\underset{\sim}{M}_i\underset{\sim}{M}'_i \mid \underset{\sim}{M}_i) = -\sum_i \underset{\sim}{M}_i\underset{\sim}{M}'_i \quad (104)$$

which means, dividing both sides by $I$, that identification asymptotically depends on the linear independence of the controls which constitute $\underset{\sim}{M}_i$. This is the same condition which appeared in both the binary choice and linear-in-means. However, if $m_{n(i)}$ is the within-group mean, then by eq. (99) $m_{n(i)}$ is a nonlinear function of $\underset{\sim}{Y}_{n(i)}$ and $F_{\underset{\sim}{X}}$.

A nonlinear relationship of this type is the key condition for global identification in the binary choice model. Extensions of the analysis for binary choice may be made to the exponential hazard model as well as other parametric cases such as the Weibull, log normal and log-logistic distributions and can further be done for cases such as right censoring or for Cox's partial likelihood approach. A formal characterization of the conditions for identification in these various cases is left for future work. In particular, we believe that analogous conditions for local

identification to those found in Theorem 7 can be developed for the current case.

### *iii.* Nonparametric approaches

#### *a.* treatment effects

Our discussion of identification has assumed that a researcher possesses prior information concerning the form of the model under study, so that estimation occurs with respect to a finite set of parameters. In our context, this information has taken the form of both the functional form for individual behavior and, where selection into neighborhoods is an issue, the rules for neighborhood self-selection. Dissatisfaction with the assumption of such strong prior information has led to a vast literature on semi- and non-parametric approaches to estimation. In the context of interactions-based models, one can think of a nonparametric approach to estimation in the context of identifying a role for neighborhood characteristics on individual behavior while making relatively weak assumptions on the functional forms describing individual behavior. In turn, one can think of this question as analogous to the nonparametric identification of treatment effects, where group influences are the "treatment" whose effect we wish to uncover.

In this section, we develop approaches to both point and interval identification of interaction effects under substantially weaker modelling assumptions than we have employed thus far. First, following Heckman (1997), we show how the assumption that neighborhood interaction effects act as a "shifted outcome effect" combined with an exclusion restriction on the determinants of neighborhood membership, can lead to identification of an interaction effect. Second, following Manski (1995) and Manski and Pepper (1998), we show how one may relax this exclusion restriction and nevertheless

obtain an upper bound on the interaction effect.

To make our analysis concrete, suppose that, following the work of Steinberg (1996) we are interested in determining whether a peer group of "brains" (denoted as group 1) versus a peer group of "nonbrains" (denoted as group 0) affects individual student performance. Observations are available from $G$ different schools, each of which contains students who are members of each such group. The variable $\xi_{i,g}$ tracks the group of individual $i$ in school $g$. The goal of the exercise is to determine the effect of membership in group 1 versus group 0 on a continuous outcome variable, $\omega_{i,g,\xi}$. Notice that we index according to both school and group. Membership in the brains groups is therefore our "treatment" and so we wish to measure the treatment effect. We let $\underset{\sim}{X}_{i,g}$ denote those observable individual variables which directly determine $\omega_{i,g,\xi}$ and $\underset{\sim}{R}_{i,g}$ denote those observable variables which determine whether $i$ is a member of group 1. We refer to the average behaviors in the two groups as $m_{g,0}$ and $m_{g,1}$ with $\underset{\sim}{m}_g = (m_{g,0}, m_{g,1})$. We have included $g$ in the subscripts so that each observation refers to both an individual and the school which he attends.

For a given individual $i$, we assume that $\omega_{i,g,\xi}$ obeys

$$\omega_{i,g,\xi} = \phi(\xi_{i,g}, \underset{\sim}{X}_{i,g}, \underset{\sim}{R}_{i,g}, \underset{\sim}{m}_g) + \epsilon_{i,g}(\xi_{i,g}) \tag{105}$$

for some function $\phi(\cdot, \cdot, \cdot, \cdot)$ where

$$\mathrm{E}(\epsilon_{i,g}(\xi_{i,g}) \mid \xi_{i,g}, \underset{\sim}{X}_{i,g}, \underset{\sim}{R}_{i,g}, \underset{\sim}{m}_g) = 0. \tag{106}$$

The identification question therefore refers to what can be learned about $\Delta_{i,g} = \omega_{i,g,1} - \omega_{i,g,0}$. Following Heckman (1997), one is typically interested in

$$\mathrm{E}(\Delta_{i,g} \mid \underset{\sim}{X}_{i,g}, \underset{\sim}{R}_{i,g}, \underset{\sim}{m}_g) = \phi(1, \underset{\sim}{X}_{i,g}, \underset{\sim}{R}_{i,g}, \underset{\sim}{m}_g) - \phi(0, \underset{\sim}{X}_{i,g}, \underset{\sim}{R}_{i,g}, \underset{\sim}{m}_g) \tag{107}$$

where the equality follows immediately from (105) and (106). This is the

expected value of the treatment for an individual with characteristics $\underset{\sim}{X}_{i,g}$ and $\underset{\sim}{R}_{i,g}$ in school $g$ and represents the object which we wish to estimate. A distinct quantity of interest is $\mathrm{E}(\Delta_{i,g} \mid \underset{\sim}{X}_{i,g}, \underset{\sim}{R}_{i,g}, \underset{\sim}{m}_g, \xi_{i,g} = 1)$ which Heckman (1997) refers to as the effect of the "treatment on the treated for persons with characteristics" $\underset{\sim}{X}_{i,g}$ and $\underset{\sim}{R}_{i,g}$. Notice that the selection problem holds because there is information about $\epsilon_{i,g}(0)$ and $\epsilon_{i,g}(1)$ when the treatment, i.e. group membership, is a choice variable.

In order to identify $\mathrm{E}(\Delta_{i,g} \mid \underset{\sim}{X}_{i,g}, \underset{\sim}{R}_{i,g}, \underset{\sim}{m}_g)$, one proceeds as follows. First, assume that the effect of group membership is additive, so that

$$\omega_{i,g,1} - \omega_{i,g,0} = k(\underset{\sim}{X}_{i,g}, \underset{\sim}{m}_g) \tag{108}$$

for some function $k(\,\cdot\,,\,\cdot\,)$ which means that

$$k(\underset{\sim}{X}_{i,g}, \underset{\sim}{m}_g) = \mathrm{E}(\Delta_{i,g} \mid \underset{\sim}{X}_{i,g}, \underset{\sim}{R}_{i,g}, \underset{\sim}{m}_g). \tag{109}$$

Eq. (108) is often referred to as a shifted outcome assumption. Notice that this is a minor generalization of Heckman (1997), although not Heckman and Robb (1985), in that we allow the $k$'s to vary with respect $\underset{\sim}{X}_{i,g}$ and $\underset{\sim}{m}_g$, which is natural if one thinks the treatment effect varies across individuals.

Next, consider what is estimable from the data. The group means $\underset{\sim}{m}_g$ are of course observable. Further, one can estimate the conditional expectations of behavior for individual given their group memberships, i. e.

$$\mathrm{E}(\omega_{i,g,0} \mid \underset{\sim}{X}_{i,g}, \underset{\sim}{R}_{i,g}, \underset{\sim}{m}_g, \xi_{i,g} = 0) \tag{110}$$

and

$$\mathrm{E}(\omega_{i,g,1} \mid \underset{\sim}{X}_{i,g}, \underset{\sim}{R}_{i,g}, \underset{\sim}{m}_g, \xi_{i,g} = 1) \tag{111}$$

The identification of an endogenous interaction effect can be thought of as requiring that one can move from these conditional expectations to $\mathrm{E}(\omega_{i,g,0} \mid \underset{\sim}{X}_{i,g}, \underset{\sim}{R}_{i,g}, \underset{\sim}{m}_g)$ and $\mathrm{E}(\omega_{i,g,1} \mid \underset{\sim}{X}_{i,g}, \underset{\sim}{R}_{i,g}, \underset{\sim}{m}_g)$. To do this, it is necessary to be somewhat more careful about the process of group formation. We therefore assume that individuals join groups at least partially on the basis of the expected behavior in the groups, and that these expected behaviors are rational. This is nothing more than the self-consistency idea we have used throughout.

We now can consider the estimation of $k(\underset{\sim}{X}_{i,g}, \underset{\sim}{m}_g)$. Letting $\mu(\xi_{i,g} \mid \underset{\sim}{X}_{i,g}, \underset{\sim}{R}_{i,g}, \underset{\sim}{m}_g)$ denote the conditional probability of group membership, it is immediate that

$$\mathrm{E}(\omega_{i,g,0} \mid \underset{\sim}{X}_{i,g}, \underset{\sim}{R}_{i,g}, \underset{\sim}{m}_g) =$$

$$\mathrm{E}(\omega_{i,g,0} \mid \underset{\sim}{X}_{i,g}, \underset{\sim}{R}_{i,g}, \underset{\sim}{m}_g, \xi_{i,g} = 0)\mu(\xi_{i,g} = 0 \mid \underset{\sim}{X}_{i,g}, \underset{\sim}{R}_{i,g}, \underset{\sim}{m}_g) +$$

$$\mathrm{E}(\omega_{i,g,0} \mid \underset{\sim}{X}_{i,g}, \underset{\sim}{R}_{i,g}, \underset{\sim}{m}_g, \xi_{i,g} = 1)\mu(\xi_{i,g} = 1 \mid \underset{\sim}{X}_{i,g}, \underset{\sim}{R}_{i,g}, \underset{\sim}{m}_g) \tag{112}$$

and

$$\mathrm{E}(\omega_{i,g,1} \mid \underset{\sim}{X}_{i,g}, \underset{\sim}{R}_{i,g}, \underset{\sim}{m}_g) =$$

$$\mathrm{E}(\omega_{i,g,1} \mid \underset{\sim}{X}_{i,g}, \underset{\sim}{R}_{i,g}, \underset{\sim}{m}_g, \xi_{i,g} = 0)\mu(\xi_{i,g} = 0 \mid \underset{\sim}{X}_{i,g}, \underset{\sim}{R}_{i,g}, \underset{\sim}{m}_g) +$$

$$\mathrm{E}(\omega_{i,g,1} \mid \underset{\sim}{X}_{i,g}, \underset{\sim}{R}_{i,g}, \underset{\sim}{m}_g, \xi_{i,g} = 1)\mu(\xi_{i,g} = 1 \mid \underset{\sim}{X}_{i,g}, \underset{\sim}{R}_{i,g}, \underset{\sim}{m}_g) \tag{113}$$

The right hand terms $\mathrm{E}(\omega_{i,g,0} \mid \underset{\sim}{X}_{i,g}, \underset{\sim}{R}_{i,g}, \underset{\sim}{m}_g, \xi_{i,g} = 1)$ and $\mathrm{E}(\omega_{i,g,1} \mid \underset{\sim}{X}_{i,g}, \underset{\sim}{R}_{i,g}, \underset{\sim}{m}_g, \xi_{i,g} = 0)$ are not observed since they refer to conditional expectations of behavior for individuals were they members of groups which they did not select into. Hence identification will only occur if some additional assumption overcomes this.

One such assumption is an exclusion restriction with respect to the variables which affect selection into groups versus variables which affect behavior once one is a member of a given the group. Formally, we need the following. For every set of pairs $\underset{\sim}{X}_{i,g}, \underset{\sim}{R}_{i,g}$ and $\underset{\sim}{X}_{i,g}, \underset{\sim}{R}'_{i,g}$

$$\mathrm{E}(\omega_{i,g,0} \mid \underset{\sim}{X}_{i,g}, \underset{\sim}{R}_{i,g}, \underset{\sim}{m}_g) = \mathrm{E}(\omega_{i,g,0} \mid \underset{\sim}{X}_{i,g}, \underset{\sim}{R}'_{i,g}, \underset{\sim}{m}_g) \tag{114}$$

and

$$\mathrm{E}(\omega_{i,g,1} \mid \underset{\sim}{X}_{i,g}, \underset{\sim}{R}_{i,g}, \underset{\sim}{m}_g) = \mathrm{E}(\omega_{i,g,1} \mid \underset{\sim}{X}_{i,g}, \underset{\sim}{R}'_{i,g}, \underset{\sim}{m}_g) \tag{115}$$

What this means is that there is a variable which affects selection but not expected behavior for each individual once that person is a group member.

Following Heckman (1997) and Manski (1995) p. 144, the shifted outcome restriction (108) and the exclusion restriction described by eqs. (114) and (115) can be combined to conclude that

$$\begin{aligned}
&\mathrm{E}(\omega_{i,g,1} \mid \underset{\sim}{X}_{i,g}, \underset{\sim}{R}_{i,g}, \underset{\sim}{m}_g, \xi_{i,g} = 1)\mu(\xi_{i,g} = 1 \mid \underset{\sim}{X}_{i,g}, \underset{\sim}{R}_{i,g}, \underset{\sim}{m}_g) + \\
&\mathrm{E}(\omega_{i,g,0} \mid \underset{\sim}{X}_{i,g}, \underset{\sim}{R}_{i,g}, \underset{\sim}{m}_g, \xi_{i,g} = 0)\mu(\xi_{i,g} = 0 \mid \underset{\sim}{X}_{i,g}, \underset{\sim}{R}_{i,g}, \underset{\sim}{m}_g) + \\
&\qquad k(\underset{\sim}{X}_{i,g}, \underset{\sim}{m}_g)\mu(\xi_{i,g} = 0 \mid \underset{\sim}{X}_{i,g}, \underset{\sim}{R}_{i,g}, \underset{\sim}{m}_g) = \\
&\mathrm{E}(\omega_{i,g,1} \mid \underset{\sim}{X}_{i,g}, \underset{\sim}{R}'_{i,g}, \underset{\sim}{m}_g, \xi_{i,g} = 1)\mu(\xi_{i,g} = 1 \mid \underset{\sim}{X}_{i,g}, \underset{\sim}{R}'_{i,g}, \underset{\sim}{m}_g) + \\
&\mathrm{E}(\omega_{i,g,0} \mid \underset{\sim}{X}_{i,g}, \underset{\sim}{R}'_{i,g} \underset{\sim}{m}_g, \xi_{i,g} = 0)\mu(\xi_{i,g} = 0 \mid \underset{\sim}{X}_{i,g}, \underset{\sim}{R}'_{i,g}, \underset{\sim}{m}_g) + \\
&\qquad k(\underset{\sim}{X}_{i,g}, \underset{\sim}{m}_g)\mu(\xi_{i,g} = 0 \mid \underset{\sim}{X}_{i,g}, \underset{\sim}{R}'_{i,g}, \underset{\sim}{m}_g).
\end{aligned} \tag{116}$$

Other than $k(\underset{\sim}{X}_{i,g}, \underset{\sim}{m}_g)$, each of the terms in this expression can be estimated

nonparametrically, and so $k(\underset{\sim}{X}_{i,g}, \underset{\sim}{m}_g)$ is identified.

**Theorem 8. Nonparametric identification of endogenous interaction effect**

In the presence of self-consistent expectations, shifted outcomes of the form eq. (108), and an exclusion restriction of the form eqs. (114) and (115), the interaction effect is identified.

Two features of this result are worth noting. First, under Theorem 8, one can estimate $\mathrm{E}(k(\underset{\sim}{X}_{i,g}, \underset{\sim}{m}_g) \mid \underset{\sim}{m}_g)$ and test for the average interaction effect in a population. Further, if one is willing to assume that

$$k(\underset{\sim}{X}_{i,g}, \underset{\sim}{m}_g) = J_g(m_{g,1} - m_{g,0}) \tag{117}$$

then the interaction parameter $J_g$ for each school may be identified. In principle, cross-school variation in $J_g$ could be employed to study the determinants of the strength of interactions.

Second, while Theorem 8 makes some progress in terms of relaxing the parametric assumptions of the interactions model, it is still strong in terms of the underlying behavioral assumptions. As stated by Heckman (1997),

"Any valid application of the method of instrumental variables for estimating these treatment effects in the case where the response to treatment varies among persons requires a behavioral assumption about how persons make their decisions about program participation. This issue cannot be settled by a statistical analysis." (pg. 449)

In our context, the treatment is the membership in group 1 rather than group 0 and the instrument is characterized by eqs. (114) and (115).

One approach to weakening the exclusion restriction on instruments is due to Manski and Pepper (1998). Following their analysis, we replace our previous

assumptions with inequality restrictions. First, we replace our assumption of a shifted outcome variable, eq. (108) with

$$\omega_{i,g,1} - \omega_{i,g,0} \leq k(\underset{\sim}{X}_{i,g}, \underset{\sim}{m}_g). \tag{118}$$

This assumption means that the effect of shifting a person with individual characteristics $\underset{\sim}{X}_{i,g}$ from group 0 to group 1 is bounded from above by $k(\underset{\sim}{X}_{i,g}, \underset{\sim}{m}_g)$. Second, we assume that a monotonic increase in the selection variables $\underset{\sim}{R}_{i,g}$ never decreases the expected outcome for an individual within a given group. Formally, if $\underset{\sim}{R}'_{i,g} \geq \underset{\sim}{R}_{i,g}$, then

$$\mathrm{E}(\omega_{i,g,\xi} \mid \underset{\sim}{X}_{i,g}, \underset{\sim}{R}_{i,g}, \underset{\sim}{m}_g) \leq \mathrm{E}(\omega_{i,g,\xi} \mid \underset{\sim}{X}_{i,g}, \underset{\sim}{R}'_{i,g}, \underset{\sim}{m}_g),\ \xi = 0,1 \tag{119}$$

This assumption relaxes eqs. (114) and (115) in that a monotonic increase from $\underset{\sim}{R}_{i,g}$ to $\underset{\sim}{R}'_{i,g}$ may have an effect on the conditional expectation of $\omega_{i,g,\xi}$, but this effect's sign must not be negative.

Under these assumptions, we have

$$\begin{aligned}
&\mathrm{E}(\omega_{i,g,1} \mid \underset{\sim}{X}_{i,g},\ \underset{\sim}{R}_{i,g}, \underset{\sim}{m}_g, \xi_{i,g} = 1)\mu(\xi_{i,g} = 1 \mid \underset{\sim}{X}_{i,g}, \underset{\sim}{R}_{i,g}, \underset{\sim}{m}_g) + \\
&\mathrm{E}(\omega_{i,g,0} \mid \underset{\sim}{X}_{i,g}, \underset{\sim}{R}_{i,g}, \underset{\sim}{m}_g, \xi_{i,g} = 0)\mu(\xi_{i,g} = 0 \mid \underset{\sim}{X}_{i,g}, \underset{\sim}{R}_{i,g}, \underset{\sim}{m}_g) + \\
&k(\underset{\sim}{X}_{i,g}, \underset{\sim}{m}_g)\mu(\xi_{i,g} = 0 \mid \underset{\sim}{X}_{i,g}, \underset{\sim}{R}_{i,g}, \underset{\sim}{m}_g) \leq \\
&\mathrm{E}(\omega_{i,g,1} \mid \underset{\sim}{X}_{i,g},\ \underset{\sim}{R}'_{i,g}, \underset{\sim}{m}_g, \xi_{i,g} = 1)\mu(\xi_{i,g} = 1 \mid \underset{\sim}{X}_{i,g}, \underset{\sim}{R}'_{i,g}, \underset{\sim}{m}_g) + \\
&\mathrm{E}(\omega_{i,g,0} \mid \underset{\sim}{X}_{i,g}, \underset{\sim}{R}'_{i,g}, \underset{\sim}{m}_g, \xi_{i,g} = 0)\mu(\xi_{i,g} = 0 \mid \underset{\sim}{X}_{i,g}, \underset{\sim}{R}'_{i,g}, \underset{\sim}{m}_g) + \\
&k(\underset{\sim}{X}_{i,g}, \underset{\sim}{m}_g)\mu(\xi_{i,g} = 0 \mid \underset{\sim}{X}_{i,g}, \underset{\sim}{R}'_{i,g}, \underset{\sim}{m}_g).
\end{aligned} \tag{120}$$

We may now consider the quantity, $Q(\underset{\sim}{X}_{i,g},\underset{\sim}{R}_{i,g},\underset{\sim}{m}_g)$ defined as

$$Q(\underset{\sim}{X}_{i,g},\underset{\sim}{R}_{i,g},\underset{\sim}{m}_g) =$$

$$\mathrm{E}(\omega_{i,g,1} \mid \underset{\sim}{X}_{i,g},\underset{\sim}{R}_{i,g},\underset{\sim}{m}_g,\xi_{i,g}=1)\mu(\xi_{i,g}=1 \mid \underset{\sim}{X}_{i,g},\underset{\sim}{R}_{i,g},\underset{\sim}{m}_g) +$$

$$\mathrm{E}(\omega_{i,g,0} \mid \underset{\sim}{X}_{i,g},\underset{\sim}{R}_{i,g},\underset{\sim}{m}_g,\xi_{i,g}=0)\mu(\xi_{i,g}=0 \mid \underset{\sim}{X}_{i,g},\underset{\sim}{R}_{i,g},\underset{\sim}{m}_g) \tag{121}$$

This term is an observable analog of the expected outcome of an individual with observed characteristics $\underset{\sim}{X}_{i,g},\underset{\sim}{R}_{i,g}$, and $\underset{\sim}{m}_g$. Inequality (118) implies that

$$Q(\underset{\sim}{X}_{i,g},\underset{\sim}{R}'_{i,g},\underset{\sim}{m}_g) + k(\underset{\sim}{X}_{i,g},\underset{\sim}{m}_g)\mu(\xi_{i,g}=0 \mid \underset{\sim}{X}_{i,g},\underset{\sim}{R}'_{i,g},\underset{\sim}{m}_g) \geq$$

$$Q(\underset{\sim}{X}_{i,g},\underset{\sim}{R}_{i,g},\underset{\sim}{m}_g) + k(\underset{\sim}{X}_{i,g},\underset{\sim}{m}_g)\mu(\xi_{i,g}=0 \mid \underset{\sim}{X}_{i,g},\underset{\sim}{R}_{i,g},\underset{\sim}{m}_g) \tag{122}$$

which may be rewritten as

$$Q(\underset{\sim}{X}_{i,g},\underset{\sim}{R}'_{i,g},\underset{\sim}{m}_g) - Q(\underset{\sim}{X}_{i,g},\underset{\sim}{R}_{i,g},\underset{\sim}{m}_g) \geq$$

$$k(\underset{\sim}{X}_{i,g},\underset{\sim}{m}_g)(\mu(\xi_{i,g}=0|\underset{\sim}{X}_{i,g},\underset{\sim}{R}_{i,g},\underset{\sim}{m}_g) - \mu(\xi_{i,g}=0|\underset{\sim}{X}_{i,g},\underset{\sim}{R}'_{i,g},\underset{\sim}{m}_g)) \tag{123}$$

So long as

$$Q(\underset{\sim}{X}_{i,g},\underset{\sim}{R}'_{i,g},\underset{\sim}{m}_g) - Q(\underset{\sim}{X}_{i,g},\underset{\sim}{R}_{i,g},\underset{\sim}{m}_g) > 0 \tag{124}$$

and

$$\mu(\xi_{i,g}=0 \mid \underset{\sim}{X}_{i,g},\underset{\sim}{R}_{i,g},\underset{\sim}{m}_g) - \mu(\xi_{i,g}=0 \mid \underset{\sim}{X}_{i,g},\underset{\sim}{R}'_{i,g},\underset{\sim}{m}_g) > 0 \tag{125}$$

one can construct an upper bound on $k(\underset{\sim}{X}_{i,g},\underset{\sim}{m}_g)$. Formulating the bound using the fact that

$$\mu(\xi_{i,g}=0 \mid \underset{\sim}{X}_{i,g},\underset{\sim}{R}_{i,g},\underset{\sim}{m}_g)-\mu(\xi_{i,g}=0 \mid \underset{\sim}{X}_{i,g},\underset{\sim}{R}'_{i,g},\underset{\sim}{m}_g)=$$

$$\mu(\xi_{i,g}=1 \mid \underset{\sim}{X}_{i,g},\underset{\sim}{R}'_{i,g},\underset{\sim}{m}_g)-\mu(\xi_{i,g}=1 \mid \underset{\sim}{X}_{i,g},\underset{\sim}{R}_{i,g},\underset{\sim}{m}_g)$$

we have Theorem 9.

**Theorem 9. Construction of upper bound on the endogenous interaction effect**

Assume that eqs. (118), (119), (124), and (125) hold. Then

$$k(\underset{\sim}{X}_{i,g},\underset{\sim}{m}_g)\leq$$

$$\frac{Q(\underset{\sim}{X}_{i,g},\underset{\sim}{R}'_{i,g},\underset{\sim}{m}_g)-Q(\underset{\sim}{X}_{i,g},\underset{\sim}{R}_{i,g},\underset{\sim}{m}_g)}{\mu(\xi_{i,g}=1 \mid \underset{\sim}{X}_{i,g},\underset{\sim}{R}'_{i,g},\underset{\sim}{m}_g)-\mu(\xi_{i,g}=1 \mid \underset{\sim}{X}_{i,g},\underset{\sim}{R}_{i,g},\underset{\sim}{m}_g)} \tag{126}$$

A weakness of this result is that when it comes to interaction effects, it is probably more interesting to obtain a lower bound, since the presence of such effects are controversial. Notice, however, that the Manski and Pepper approach does suggest a way of constructing such a lower bound. In order to do so, one would need to find a variable which possesses the features that an increase (decrease) in its level would 1) both increase (decrease) the probability of selection into group 1 and 2) decrease (increase) the expected outcome for an individual conditional on the variable. Introspection suggest that it may be difficult to find such a variable, although there may be contexts where it holds.

In contrast, the assumption that one can find a variable in which both effects move in the same direction seems relatively plausible. For example, Manski and Pepper consider the question of how to bound the effect of the returns

to schooling, using SAT as an instrument. It seems natural in this case to assume both that higher SAT's make additional schooling more likely and higher SAT's do not reduce the benefit of additional schooling if chosen.

The self-consistency conditions

$$m_{\xi,g} = \int \mathrm{E}(\omega_{i,g,\xi} \mid \xi, \underset{\sim}{X}, \underset{\sim}{R}, \underset{\sim}{m}_g) dF_{\underset{\sim}{X},\underset{\sim}{R},g}, \ \xi = 0,1 \tag{127}$$

(where following our previous convention, $dF_{\underset{\sim}{X},\underset{\sim}{R},g}$ denotes the joint distribution of $\underset{\sim}{X}$ and $\underset{\sim}{R}$ in school $g$), play an essential role in determining the quality of the bound in (126). To see this, consider the extreme case where all individuals within a school $g$ have identical values of $\underset{\sim}{X}_{i,g}$ and $\underset{\sim}{R}_{i,g}$, then the bound is undefined, since the numerator and denominator of (125) will each equal 0. Alternatively, suppose that the distribution functions of individual characteristics are identical across schools, so $dF_{\underset{\sim}{X},\underset{\sim}{R},g} = dF_{\underset{\sim}{X},\underset{\sim}{R},g'}$. If the solutions to the self-consistency conditions (127) are unique, then this implies expected average outcomes must also be identical, i.e. $\underset{\sim}{m}_g = \underset{\sim}{m}_{g'}$. If $k(\underset{\sim}{X}_{i,g}, \underset{\sim}{m}_g) = k(\underset{\sim}{m}_g)$, so that the bound does not depend on individual characteristics, then one can use cross-school information in that $k(\underset{\sim}{m}_g)$ must be bounded by the $inf$ of the upper bounds computed for each school in isolation. This type of argumentation seems a valuable area for future research.

At a minimum, this discussion illustrates two points. First, semi- and non-parametric approaches to inference can be adapted to achieve either point or interval identification of interaction effects. Second, the conditions required for identification require careful consideration of the underlying socioeconomic theories under analysis in order to identify appropriate instruments.

***b.* duration data**

Identification can also be considered for nonparametric approaches to duration data. As compellingly demonstrated by Heckman and Singer (1984b), errors in the assumed form of the hazard function and in the "mixing distribution" (by which they mean the distribution of unobservables) can lead to wildly misleading estimates. These problems of course are also relevant when interaction effects may be present. As far as we know, the extension of the methods studied by Heckman and Singer and subsequent authors to models with interactions has yet to be studied and it is beyond the scope of this paper to do so. However, we do sketch a slight extension to one approach to nonparametric identification in duration models, due to Elbers and Ridder (1982), in order to illustrate how such argumentation can in principle proceed; the reader is advised to see Heckman and Singer (1984c) for an evaluation of alternative conditions for nonparametric identification in this context.

Following Elbers and Ridder, suppose that the hazard function may be written as

$$\lambda_i(t) = a(t)h(\underset{\sim}{M}_i)\nu_i \tag{128}$$

Relative to our original treatment of hazards, this incorporates two additional terms: $a(t)$ which allows for duration dependence, and $\nu_i$ which allows for unobserved heterogeneity. The $\nu_i$'s are assumed to be drawn from a common distribution $F_\nu(\cdot)$ with associated density $f_\nu(\cdot)$. Elbers and Ridder show that subject to appropriate regularity conditions, if $F(t \mid \underset{\sim}{M}_i)$ is nondefective (which means that all spells are completed), then it is possible to identify $a(\cdot)$, $h(\cdot)$, and $f_\nu(\cdot)$. They do this as follows.

Define the conditional survivor function for individual $i$ as

$$S(t, \underset{\sim}{M}_i, \nu_i) = exp(\int_0^t a(r)h(\underset{\sim}{M}_i)\nu_i dr) = exp(A(t)h(\underset{\sim}{M}_i)\nu_i) \tag{129}$$

where $A(t) = \int_0^t a(r)dr$ and the conditional survivor function

$$S(t,\underset{\sim}{M}_i) = \int S(t,\underset{\sim}{M}_i,\nu)dF_\nu = \int exp(\tau\nu)dF_\nu \tag{130}$$

where $\tau = A(t)h(\underset{\sim}{M}_i)$. The last term in (130) indicates how the conditional survivor function is the LaPlace transform of $dF_\nu$. The analyst is assumed to observe a family of nondefective distribution functions $G(t,\underset{\sim}{M}_i) = 1 - S(t,\underset{\sim}{M}_i)$ from which he wishes to recover $a(\cdot)$, $h(\cdot)$, and $f_\nu(\cdot)$. Elbers and Ridder assume that 1) $v$ is nonnegative with mean 1, 2) $\underset{\sim}{M}_i$ lies in an open set in the $k$-dimensional reals for some $k$, and 3) $h(\cdot)$ is defined on this open set and is nonnegative, differentiable, and nonconstant on the set.

Working through the proof that these conditions allow for identification, reveals the following. First, if one differentiates $G(t,\underset{\sim}{M}_i)$ with respect to $t$, $h(\underset{\sim}{M}_i)$ may be recovered regardless of whether a self-consistency condition like eq. (99) holds when $m_{n(i)}$ is a component of $\underset{\sim}{M}_i$. Second, Elbers and Ridder exploit the LaPlace transform relationship to obtain a differential equation by assuming, without loss of generality, that $\underset{\sim}{M}_i$ is one-dimensional. This argument requires differentiability and nonconstancy of $h(\underset{\sim}{M}_i)$.

In order to generalize this step to allow for endogenous interactions with a self-consistency condition such as eq. (99), the reference group for each individual $i$ must be broad enough to allow differentiability with respect to a nontrivial subvector of $\underset{\sim}{M}_i$. Specifically, one needs to be able to vary $\underset{\sim}{X}_i$ and $\underset{\sim}{Y}_{n(i)}$ without $m_{n(i)}$ varying. For example, suppose that within each neighborhood, all $\underset{\sim}{X}_i$'s are identical. In this case, the self-consistent average choice level is

$$m_{n(i)} = \mathrm{E}(t \mid \underset{\sim}{X}_i,\underset{\sim}{Y}_{n(i)},m_{n(i)}) = \int t dG(t,\underset{\sim}{X}_i,\underset{\sim}{Y}_{n(i)},m_{n(i)}). \tag{131}$$

Generically, eq. (131) will have only a finite number of self-consistent solutions (if such solutions exist). Therefore, in this case, $(\underset{\sim}{X}_i,\underset{\sim}{Y}_{n(i)})$ cannot be varied independently of $m_{n(i)}$ and it is not obvious how to adapt the Elbers and Ridder (1982) proof (pg. 405) to this case. Hence for the case of arbitrarily fine-grained

reference groups, identification is currently problematic.

In contrast, suppose that individuals reference on a coarse group $G$. In this case the self-consistency condition is

$$m_{n(i)} = \int_X \mathrm{E}(t \mid X, \underset{\sim}{Y}_{n(i)}, m_{n(i)}) dF_X = H(m_{n(i)}) \tag{132}$$

where $dF_X$ is the distribution of $X$'s within $n(i)$. In this case, it is possible to locate a nontrivial set of sufficient conditions on $dG(t, \underset{\sim}{X}_i, \underset{\sim}{Y}_{n(i)}, m_{n(i)})$ such that the hazard function is differentiable with respect to $(\underset{\sim}{X}_i, \underset{\sim}{Y}_{n(i)})$ on the self-consistent solutions defined by eq. (132). This appears to be sufficient to extend Elbers and Ridder's identification argument to the case of interactions.

## 5. Sampling properties

In this section, we develop some asymptotics for the parameter estimates for interactions-based models and consider the effects on such estimates of omitted variables. The sampling properties for data generated by interactions-based models are no different from that associated with standard discrete choice and linear regression models. The critical property which one needs to verify is that the behavioral data obey the standard limits theorem necessary for asymptotics when there is sufficient dependence across observations to induce multiple equilibria. Similarly, the effects of omitted variables mirror results found in other contexts. We therefore focus on the binary choice case.

### *i.* laws of large numbers

Despite the dependence introduced by interactions, the data generated by the noncooperative version of the binary choice model with interactions generates

a law of large numbers. Brock and Durlauf (1995) showed this for the special case where $h_i = h \ \forall \ i$; it is straightforward (cf. Ash (1972) pg. 234) to extend this result to non-identically distributed choices.

**Theorem 10. Law of large numbers for realized average choice levels in noncooperative version of the binary choice model with interactions**

Suppose that a population of agents holds a common belief that the expected value of the average population choice is $m^*$, where $m^*$ is a solution to eq. (22). Then a weak law of large numbers will apply as $I$ becomes arbitrarily large such that

$$lim \ \bar{\omega}_I \Rightarrow_w m^*. \tag{133}$$

*ii.* **naive estimator**

The "naive" estimator whose identification properties we have analyzed does not introduce any new econometric issues with respect to asymptotic normality. Theorem 11 is standard; necessary conditions for it are given, for example, in McFadden (1984), p. 1399. The specific conditions we cite are found in Amemiya (1985) pg. 270.[5]

**Theorem 11. Consistency and asymptotic normality of naive estimates in the binary choice model with global interactions**

Let $b = (k,c,d,J)'$ and $\underset{\sim}{M}_i = (1,\underset{\sim}{X}_i',\underset{\sim}{Y}_{n(i)}',m_{n(i)}^e)'$. If the binary choice model with

[5]In Amemiya (1985), it is also assumed that the $\underset{\sim}{M}_i$ elements are uniformly bounded when asymptotic normality is proved, which contradicts our identification assumption that the $\underset{\sim}{Y}_{n(i)}$'s are unbounded. However, as Amemiya points out, the boundedness assumption can be dispensed with.

interactions is globally identified, and if

*i.* $b$ lies in an open, bounded subset of $R^{r+s+2}$.

*ii.* $lim_{I \Rightarrow \infty} I^{-1} \sum_{i \in I} \underset{\sim}{M}_i \underset{\sim}{M}_i'$ is a finite nonsingular matrix.

*iii.* The empirical distribution function of $\underset{\sim}{M}_i$ converges to a distribution function.

Then, the maximum likelihood estimates $\hat{b}_I$ of the binary choice model with global interactions are consistent and asymptotically normal with limiting behavior

$$I^{1/2}(\hat{b}_I - b) \Rightarrow_w N(0, \mathfrak{I}^{-1}) \tag{134}$$

where

$$\mathfrak{I} = lim_{I \Rightarrow \infty} I^{-1} \sum_i \frac{exp(\hat{b}' \underset{\sim}{M}_i)}{(1 + exp(\hat{b}' \underset{\sim}{M}_i))^2} \underset{\sim}{M}_i \underset{\sim}{M}_i' \tag{135}$$

($\mathfrak{I}$ is of course the suitably normalized information matrix of the likelihood function and is consistently estimable for this model.)

### *iii.* asymptotics for data generated by social planner

Models which incorporate realized contemporaneous interactions between individuals introduce several mathematical complexities relative to standard econometric models. As noted before, this occurs because of the quadratic terms which appear in the likelihood. The following theorem is proved in the Mathematical Appendix; unlike Theorem 10 it does not apply to the case of

heterogeneous $h_i$'s, although results in Amaro de Matos and Perez (1991) suggest this can be done.

**Theorem 12. Large economy limit for realized average choice levels in social planner's version of the binary choice model with interactions**

Suppose that the vector of choices in a population is determined by a social planner with preferences consistent with eq. (44). The sample mean of these choices will converge weakly such that

$$lim\ \overline{\omega}_I \Rightarrow_w m^* \tag{136}$$

where $m^*$ is the solution to $m^* = tanh(\beta h + \beta J m^*)$ with the same sign as $h$.

Unfortunately, maximum likelihood estimation has yet to be developed for data generated by a social planner problem of the type we have studied. While techniques developed in Amaro de Matos and Perez (1991), Brock (1993), and Ellis (1985) all suggest that the development of these asymptotics is feasible, the argument seems sufficiently complicated that we are not comfortable making a conjecture on the asymptotic distribution of the estimator. In the Mathematical Appendix, we provide some initial discussion of these issues to illustrate how such a theory could be developed.

*iv.* **unobserved variables**

Perhaps the most serious criticism made of efforts to identify interaction effects is the difficulty in identifying interaction effects in the presence of unobserved individual or group characteristics. This is true because the main

groupings for which interactions are conjectured to exist, neighborhoods, schools, firms, etc., the neighborhoods are endogenously determined. Presumably, neighborhood contextual and endogenous characteristics influence individual choices as to neighborhood membership. Hence, it seems very likely that omitted variables which influence individual behavior once that person is a member of a neighborhood will also be correlated with the various group effects which are captured in a statistical model. This point is distinct from the self-selection issues which are discussed above.

In particular, we are interested in determining how omitted variables will affect inferences concerning $J$. We do this following a maximum likelihood approach due to Cameron and Heckman (1998). This approach is straightforward to describe in sample, rather than population terms which is why we place it here.

In our framework, assume that the binary choices $\omega_i$ are coded 0,1 and are generated by the probability model

$$\mu(\omega_i = 1 \mid \underset{\sim}{X}_{i,o}, \underset{\sim}{Y}_{i,o}, \underset{\sim}{X}_{i,u}, \underset{\sim}{Y}_{i,u}, m^e_{n(i)}) =$$

$$F_\epsilon(k + c'_o \underset{\sim}{X}_{i,o} + d'_o \underset{\sim}{Y}_{i,o} + c'_u \underset{\sim}{X}_{i,u} + d'_u \underset{\sim}{Y}_{i,u} + J m^e_{n(i)}) \tag{137}$$

where subscripts $o$ and $u$ refer to observed and unobserved variables respectively. Let $\Theta' = (k, c'_o, d'_o, J)$, $\underset{\sim}{Z}'_i = (1, \underset{\sim}{X}_{i,o}, \underset{\sim}{Y}_{i,o} m^e_{n(i)})$ and $\eta_i = c'_u \underset{\sim}{X}_{i,u} + d'_u \underset{\sim}{Y}_{i,u}$. This means the true probability structure can be rewritten as

$$\mu(\omega_i = 1 \mid \underset{\sim}{X}_{i,o}, \underset{\sim}{Y}_{i,o}, \underset{\sim}{X}_{i,u}, \underset{\sim}{Y}_{i,u}, m^e_{n(i)}) = F_\epsilon(\Theta' \underset{\sim}{Z}_i + \tau \eta_i) \tag{138}$$

which produces a likelihood function of the form

$$L = I^{-1} \sum_i (\omega_i log F_\epsilon(\Theta' \underset{\sim}{Z}_i + \tau \eta_i) \ + (1 - \omega_i) log(1 - F_\epsilon(\Theta' \underset{\sim}{Z}_i + \tau \eta_i)) \tag{139}$$

The derivatives of the likelihood function (139) may be written as

$$L_{\Theta} = I^{-1} \sum_i \underset{\sim}{Z}_i (\omega_i - F_\epsilon(\Theta' \underset{\sim}{Z}_i + \tau \eta_i)) \tag{140}$$

$$L_{\Theta,\Theta} = -I^{-1} \sum_i f_\epsilon(\Theta' \underset{\sim}{Z}_i + \tau \eta_i) \underset{\sim}{Z}_i \underset{\sim}{Z}_i' \tag{141}$$

$$L_{\Theta,\tau} = -I^{-1} \sum_i f_\epsilon(\Theta' \underset{\sim}{Z}_i + \tau \eta_i) \underset{\sim}{Z}_i \eta_i \tag{142}$$

where $\tau = 0$ is the case where there are no unobservables and $\tau = 1$ is the case where there are unobservables. The maximum likelihood estimate of $\Theta$ must obey

$$L_{\Theta}(\Theta(\tau),\tau) = 0 \tag{143}$$

where we have written the likelihood as a function of the unknown parameters $\Theta$ and have allowed the estimate of $\Theta$ to depend on $\tau$. Further,

$$L_{\Theta,\Theta}(\Theta(\tau),\tau) \frac{d\Theta}{d\tau} + L_{\Theta,\tau}(\Theta(\tau),\tau) = 0 \tag{144}$$

which implies that

$$\frac{d\Theta}{d\tau} = -L_{\Theta,\Theta}(\Theta(\tau),\tau)^{-1} L_{\Theta,\tau}(\Theta(\tau),\tau). \tag{145}$$

Integrating both sides of this expression produces

$$\Theta(0) - \Theta(1) = \int_0^1 L_{\Theta,\Theta}(\Theta(\tau),\tau)^{-1} L_{\Theta,\tau}(\Theta(\tau),\tau) d\tau \tag{146}$$

This difference describes the effect of misspecification since, as noted above, $\tau = 0$ corresponds to the case of no unobservables whereas $\tau = 1$ corresponds to the case with unobservables as we have formulated them.

In general, one cannot determine the sign of $\Theta(0) - \Theta(1)$; this is not

surprising since it is known in contexts such as measurement error that unambiguous statements about directions of bias cannot be made. However, as recognized by Bretagnolle and Huber-Carol (1988), one can determine the sign of this bias in special cases. For example, if the elements of $L_{\Theta,\Theta}$ are all negative and that the elements of $L_{\Theta,\tau}$ are all positive, then the coefficients in the misspecified model are all biased upwards.

Using the formula (146), one can compute the bias associated with the parameter $J$. For the case where the vector $\underset{\sim}{X}_i$ is replaced with a scalar $x_i$ the vector $\underset{\sim}{Y}_{n(i)}$ is replaced with a scalar $y_{n(i)}$, one can compute

$$J(0) - J(1) =$$

$$\int_0^1 ((L_{\Theta,\Theta}^{-1})_{4,1}(I^{-1}\sum_i f_\epsilon(\Theta'\underset{\sim}{Z}_i + \tau\eta_i)\eta_i)d\tau +$$

$$\int_0^1 ((L_{\Theta,\Theta}^{-1})_{4,2}(I^{-1}\sum_i f_\epsilon(\Theta'\underset{\sim}{Z}_i + \tau\eta_i)x_i\eta_i)d\tau +$$

$$\int_0^1 (L_{\Theta,\Theta}^{-1})_{4,3}(I^{-1}\sum_i f_\epsilon(\Theta'\underset{\sim}{Z}_i + \tau\eta_i)y_{n(i)}\eta_i)d\tau +$$

$$\int_0^1 (L_{\Theta,\Theta}^{-1})_{4,4}(I^{-1}\sum_i f_\epsilon(\Theta'\underset{\sim}{Z}_i + \tau\eta_i)m_{n(i)}\eta_i)d\tau \qquad (147)$$

where $(L_{\Theta,\Theta}^{-1})_{i,j}$ is the $i,j$'th element of $L_{\Theta,\Theta}^{-1}$ and we impose self-consistency of beliefs, (i.e. substituting $m_{n(i)}$ for $m_{n(i)}^e$). The term $(L_{\Theta,\Theta}^{-1})_{4,4}$ is positive whereas the other inverse elements of these integrals are of ambiguous sign. This is only sense in which one might say there is a presumption that estimates of interaction effects are biased towards finding them because of omitted variables.

Cameron and Heckman (1998) show how the Heckman and Singer (1984b) nonparametric likelihood estimator can be used to estimate the distribution of the unobserved $\eta_i$'s and thereby compute unbiased estimates of the parameters of the observables. They make a compelling argument that the production of "heterogeneity-corrected estimates" is essential in conducting assessments of

policy experiments. We are currently pursuing the development of this idea to produce estimates of interaction effects which are robust to omitted individual and group characteristics.

## 6. Statistical analysis with grouped data

In this section, we explore some of the approaches to identifying interactions which have been developed for aggregated data. Our discussion so far has assumed that individual level observations are available to the researcher. In contexts such as economic growth or crime rates, it is often the case that only group-level data is available. As a result, there has been a distinct literature which deals with uncovering interactions from aggregated data series.

### *i.* differences in cross-group behavior

One approach to the identification of interactions from group level data is due to Glaeser, Sacerdote and Scheinkman (1996) and extended in Glaeser and Scheinkman (1998). The basic insight of this work focuses on the implications of interactions for the distribution of cross-group differences in choices. Consider a collection of groups, $N_1 ... N_I$, each of which has $n$ members. In each of the groups, individuals face a binary choice. If the individuals within each group are identical, and their choices are independent of one another, then sample means of choices within each group will scale according to the law of large numbers. Supposing that the probability of choosing 1 is $p$, then the variance of the sample average for each of the two groups is $n^{-1}p(1-p)$. This means that the cross-group variance converges weakly to zero at rate $n^{-1}$. Observations that the cross-group variance scales at a slower rate, i.e. that the cross-group differences in average choice vary too much to be consistent with the sample variance under the

null hypothesis of independent and identically distributed choices, is taken as evidence of social interactions.

Glaeser, Sacerdote, and Scheinkman (1996) apply their analysis to the study of cross-city crime rates. They find that even after controlling for city-specific socioeconomic variables, there are cross-city differences in crime rates which are far greater than would be consistent with individuals making independent choices within cities and conclude that this evidence is strongly supportive of an interactions approach.

*ii.* **spatial patterns**

Topa (1997) has attempted to identify and measure interactions through the use of spatial data. The basic idea of this work is to take seriously the idea that geographic proximity is a proxy for social proximity. Topa does this by considering the relationship between unemployment rates in census tracts in Chicago. Since census tracts typically vary in size between 2,000 to 8,000 residents, these units would seem to be good proxies for neighborhoods. Topa further assumes that the social distance between any two adjacent tracts is 1, the social distance between a tract and another tract that can be reached by travelling through a single other tract is 2, etc. Using these assumptions, he formulates the determinants of unemployment in a given tract $n$ at time $t$, $\bar{\omega}_{n,t}$ as

$$\bar{\omega}_{n,t} = \psi(c' \underset{\sim}{\bar{X}}_{n,t} + J' \underset{\sim}{\bar{\omega}}_{n,D,t} + \epsilon_{i,t}) \tag{148}$$

where $\underset{\sim}{\bar{X}}_{n,t}$ denotes census tract averages of a set of individual characteristics, $\underset{\sim}{\bar{\omega}}_{n,D,t}$ is a vector of average unemployment rates for tracts at social distances 1,2,...$D$ away from tract $n$ and $\psi(\cdot)$ is a nonlinear function generated by the stochastic model (a contact process) used to motivate the econometrics. Topa estimates this model using indirect inference methods and finds evidence of

interaction effects in the sense of a statistically significant $J$ vector.

Conley and Topa (1999) extend this analysis by attempting to identify what role different measures of distance play in explaining these spatial correlations. In particular, they construct measures of neighborhood distance based on 1) physical distance, 2) travel time, 3), ethnicity and 4) occupation. Their results suggest that physical and occupational distance explain residual correlations across unemployment rates within census tracts once intra-tract characteristics have been controlled for. Akerlof (1997) demonstrates the theoretical importance of integrating social distance into economic analysis and provides a range of interesting potential applications; the Conley and Topa work should help produce empirical measures of various types of social distance.

### *iii.* ecological inference

In the political science literature, there have been some efforts to identify group effects under the rubric of what is known as the ecological inference problem. In the basic version of this problem, a researcher possesses data on the number of whites and African Americans in each of a set of $I$ neighborhoods, as well as the number of votes received by a white and an African American candidate in the same neighborhoods. The researcher's goal is to determine the relationship between racial composition of a neighborhood and the distribution of votes by race. Since the researcher is attempting to infer individual behavior from aggregate statistics, the inference is referred to as "ecological."

Ecological inference has generated a literature which has recently begun to grow (Goodman (1953), Freedman, Klein, Sacks, Smythe, and Everett (1991), King (1997), Freedman, Klein, Ostman, Roberts (1998)). We follow the exposition of the last paper. Letting $r_i$ denote the percentage of African Americans in a neighborhood and $v_i$ as the percentage of votes accrued by an

African American candidate, the standard ecological regression is

$$v_i = pr_i + q(1 - r_i) + \epsilon_i \tag{149}$$

where $p$ is the probability with which African Americans vote for an African American candidate and $q$ is the probability with which whites vote for an African American candidate. This equation is estimable by ordinary least squares. Alternative approaches to ecological inference typically modify this equation. For example, King (1997) proposes treating the racial voting propensities as neighborhood-specific draws from a common distribution rather than as constants.

From the perspective of the sorts of data sets and models of interest to economists, we suspect that ecological inference as it has been developed is of limited interest. The formulation of the regressions fails to correspond in a natural way to the aggregate of individual decisions into group behavioral percentages in a way consistent with a choice-theoretic framework. Indeed, the consistency of aggregated voting behavior with more than one behavioral model is precisely the basis on which Freedman, Klein, Sacks, Smythe, and Everett (1991) argued that evidence of differences in $p$ and $q$ could not be interpreted in terms of underlying differences in behavior between white and African American voters. As far as we know, no one has yet shown that the statistical tools in the ecological inference literature can complement other techniques for the recovery of socioeconomic structure. However, Cross and Manski (1999) suggest new directions along these lines which may both clarify what structural mechanisms can be revealed by aggregate data as well as show how ecological regression relates to omitted variables problems in econometrics.

## 7. Evidence

In this section, we survey some of the evidence which has been adduced to

detect the presence of and to measure the magnitude of interactions. We divide the empirical literature into two parts. The first part assumes that the regression of individual outcomes on individual and group level variables represents a correctly specified model. In particular, the analysis assumes that there are no omitted variables which will generate coefficient inconsistency. The second approach accounts for the possibility of such omitted variables and explores ways to correct for inconsistency either through choice of data sets or econometric techniques.

*i.* **analyses under assumption of correct specification**

In this subsection we review some prominent empirical analyses of interactions.

**1. Neighborhood effects in youth and adult outcomes**

Perhaps the most widely empirically studied area of interactions concerns the effects on adults of the neighborhoods in which they grew up. The typical analysis of this type computes a regression of the form

$$\omega_{i,t+1} = a + c' \underset{\sim}{X}_{i,t} + d' \underset{\sim}{Y}_{n(i),t} + \epsilon_{i,t+1} \tag{150}$$

where, as before individual family characteristics and neighborhood characteristics are denoted by $\underset{\sim}{X}_{i,t}$ and $\underset{\sim}{Y}_{n(i),t}$ respectively and $E(\epsilon_{i,t+1} \mid \underset{\sim}{X}_{i,t}, \underset{\sim}{Y}_{n(i),t}) = 0$. Acceptance of the null hypothesis that $d' = \underset{\sim}{0}$ is interpreted as acceptance of the null that no interaction effects exist. Examples of this type of regression include Brooks-Gunn, Duncan, Klebanov, and Sealand (1993), Corcoran, Gordon, Laren, and Solon (1992), Rivkin (1997), and Zax and Rees (1998). These studies

typically find some combinations of $\underset{\sim}{Y}_{n(i),t}$ which are statistically significant, although there seems to be no consensus on which of these contextual effects are most robust. A useful extension of this work would be an analysis which explicitly attempted to identify robust neighborhood and individual controls, using techniques such as Leamer's (1983) extreme bounds analysis or Bayesian model averaging of the type advocated by Raftery (1994) and Raftery, Madigan, and Hoeting (1997). While these procedures do not give a definitive solution to the problem of model uncertainty, they are nevertheless invaluable in clarifying dimensions along which arbitrary model assumptions (in this case choice of control variables) matters.

Several complementary strands exist to this class of empirical research. In one approach, the importance of neighborhood-level interactions effects on inequality is evaluated by assessing the effects on inequality measures of different sorting rules. This idea is originally due to Kremer (1997); nonlinear alternatives to Kremer's original analysis have been explored by Ioannides (1997b). In a second approach, the notion of neighborhoods has shifted from geographic proximity to membership in an ethnic group. Evidence of ethnic group effects has been found by Borjas (1992,1995) and Bertrand, Luttmer, and Mullainathan (1998). Similarly, Cutler and Glaeser (1997) illustrates how segregation adversely affects a number of socioeconomic outcomes for African Americans. A third strategy has been employed by Ioannides (1999) who shows how house spatial relations in price dynamics implicitly reveal neighborhood effects. In yet a fourth approach, Solon, Page, and Duncan (1999) use within-neighborhood correlations to overcome measurement problems associated with what neighborhood attributes actually matter for interactions. They find that once various family background variables are controlled for, within-neighborhood correlations in educational attainment are low.

Finally, there is a distinct literature on the relationship between interactions and efficient and/or equilibrium sorting. Becker (1973) and Sattinger (1975) are standard references; see LeGros and Newman (1997) for the state-of-

the-art. In addition, equilibrium sorting has been studied in many contexts using Tiebout type arguments. Recent contributions include Epple and Romer (1991) and Fernandez and Rogerson (1996) whose models are directly germane to the study of inequality. In an important paper, Epple and Sieg (1999) show how to econometrically implement models of this type. Our belief is that these types of models should be further employed to provide complementary insights to the main body of literature on interactions, as the strength of interaction effects should presumably be at least partially revealed by the the neighborhood choices of individuals.

### *ii.* analyses which are robust to unobserved correlated heterogeneity

From the perspective of empirical analysis, the main issue which has concerned researchers is the problem of spurious identification of interaction effects due to the likelihood of correlated unobservables existing among individuals in endogenously determined groups.

#### *a.* matching

One approach to dealing with the possible unobserved correlated heterogeneity has attempted to identify environments which allow one to match populations subjected to different influences in order to assess the effect of changes in group membership. The most prominent type of matching study falls under the rubric of "natural experiments." By natural experiments, we refer to cases where interaction effects are identified by studying cases where some individuals that would normally be members of one group are moved to another through an exogenous intervention of some type. Those who are moved may be thought of as receiving a treatment, whereas those who remain may be thought of as a control group. While intuitively appealing, there are in fact many subtleties in analyzing

data of this type. Heckman and Smith (1995), Heckman (1996,1997), Heckman, Ichimura, Smith and Todd, (1998), Heckman, Ichimura, and Todd (1998) provide a wide ranging analysis of the salient issues. Hence an important future exercise is the reconsideration of some of these empirical studies in light of these recent econometric developments.

Among the most prominent examples of natural experiments of this type, we would list:

**1. Gautreaux Assisted Housing Program.**

In 1966, the Chicago Housing Authority was sued for discrimination by public housing residents on the grounds that both the location of public housing sites and the allocation of slots in these sites intentionally placed minorities in isolated inner city neighborhoods. In an agreement worked out between the plaintiffs and defendants, known as the Gautreaux Assisted Housing Program, housing subsidies and placement services were established for public housing residents throughout Chicago. Rosenbaum (1995) and Yinger (1995) provide reviews of the details of the Gautreaux program. For the purposes of studying interactions, several points of these features are important. The number of participants each year was fixed and so, due to oversubscription, actual participants, after some screening, were randomly selected. Families who applied for assistance were randomly given a single option of moving to another part of Chicago or to moving to a suburb. (Families who declined the offered option were placed back in the pool of eligible families from which recipients of aid were drawn.)

A series of papers, Rosenbaum and Popkin (1991), Popkin, Rosenbaum, and Meaden (1993), and Rosenbaum (1995), has analyzed the results of surveys of Gautreaux program participants in order to identify the effects of the differences between the urban and suburban environments on various socioeconomic

measures.

While they are an important source of information on interactions effects, it is important to recognize that the Gautreaux data are not ideal for this purpose. Applicants to the program were dropped who either had poor rent paying histories or who failed a home inspection to determine whether they had mistreated their public housing. This prescreening eliminated approximately 30% of the program's applicants (Rosenbaum (1995)). Further, the survey efforts conducted by Rosenbaum and coauthors exhibit some sample selection problems. In particular, those families who moved to suburbs and then returned to Chicago could not be identified. Hence, the evidence of neighborhood effects obtained from Gautreaux is, while informative, not decisive. That being said, recent work such as Rosenbaum, DeLuca, and Miller (1999), by linking Gautreaux interview data to administrative data, should be able to partially address these concerns.

The Gautreaux program also illustrates the difficulty of identifying policy effects as well as a limitation in the utility of the naive estimator in predicting the effects of changes in interaction groups. Suppose that Gautreaux families are described by a linear-in-means model of the type:

$$\omega_i = d' \underset{\sim}{Y}_{n(i)} + J m^e_{n(i)} + \epsilon_i \tag{151}$$

Suppose that one new family is moved from the inner city to a suburb. In this case, the family's presence in the new location will have no effect on either $\underset{\sim}{Y}_{n(i)}$ or $m^e_{n(i)}$ (and equivalently $m_{n(i)}$) in the new location and there will be no other Gautreaux families to reference on. In this case the knowledge of the parameters $d'$ and $J$ (which can be consistently estimated) and the neighborhood variables $\underset{\sim}{Y}_{n(i)}$ and $m^e_{n(i)}$ in the old and new locations of residence will be sufficient to predict the effect on the family of the move.

However, suppose that the Gautreaux program is expanded to the extent that clusters of families are moved from an inner city to the new neighborhood. In this case, the appropriate model is

$$\omega_i = d' \underset{\sim}{Y}_{n(i)} + J m^e_{n(i)} + d'_{G(i)} \underset{\sim}{Y}_{n(i),G(i)} + J_{G(i)} m^e_{n(i),G(i)} + \epsilon_i \qquad (152)$$

Here, $G(i)$ denotes Gautreaux families in the neighborhood, so that for example, $m^e_{n(i),G(i)}$ denotes the mean behavior of Gautreaux families in a community. Predictions of the effect of a move of a cluster of families must therefore incorporate both the effects of the move on the mean for the neighborhood as a whole as well as the possibility that the Gautreaux families will represent a subgroup within the neighborhood which induces separate interactions. This means that the move of a cluster may be subject to social multipliers of the type we have described. At a minimum, the naive estimator is no longer useful for policy and prediction analysis. The analog of the self-consistency equation (21) must now be estimated along with the individual level equation (152) in order to permit predictions of the outcomes of cluster moves.

## 2. Moving to Opportunity Demonstration

The Moving to Opportunity Demonstration is an ongoing experimental demonstration being conducted by the Department of Housing and Urban Development to evaluate the effects of moving low-income families out of high-poverty neighborhoods; a detailed discussion of the program appears in Goering (1996). The demonstration randomly assigned a set of low income families normally eligible for Section 8 housing assistance vouchers to one of three groups: 1) those eligible for housing vouchers which are only usable in census tracts with less than 10% poverty, 2) those eligible for regular Section 8 vouchers with no locational restrictions, and 3) a group whose assistance is only based on residence in a public housing project. The demonstration is being conducted in 5 metropolitan areas: Baltimore, Boston, Chicago, Los Angeles, and New York City. One motivation of the demonstration was a desire to address some of the self-selection problems associated with data from the Gautreaux Program. That

being said, it is unclear at this stage to what extent self-selection is better controlled for here than in Gautreaux, given the voluntary nature of participation in the MTO demonstrations.

Preliminary results on the various experiments are becoming available. Ladd and Ludwig (1998) report evidence that those families in Baltimore that moved out of low income census tracts achieved access to superior schools as measured by a range of criteria. However, they find little evidence that the value added of these schools for the children in these families is higher than the schools used by families in the comparison and control groups. For the Boston demonstration, Katz, Kling, and Liebman (1998) also find evidence that the MTO program has been successful in generating relocation of families, this time defined as movements out of low poverty neighborhoods. They also find that children in both types of families eligible for vouchers exhibited substantially higher test scores as well as lower incidences of behavioral problems.

### 3. Milwaukee School Voucher Program

In 1990, Wisconsin implemented the nation's first public school voucher program. In essence, this program made available school vouchers equal to the average per pupil expenditure by public schools in Milwaukee. Applicants to the program were required to fulfill several criteria. Oversubscription to the program has meant that a random subset of eligible applicants have actually been able to participate in the program. Eligibility for the program was restricted by two criteria: 1) a family could not have an annual income which exceeded 1.75 times the poverty line and 2) the student to receive the voucher could not have previously been enrolled in a private school in the year prior to the use of the voucher.

As the number of applicants greatly exceeded the number of available vouchers, the randomness of the selection process meant that there existed two groups of students, namely those who did or did not receive vouchers for private

schools, whose subsequent performance could be compared. Rouse (1998a,b) and Witte (1997) have both studied this question. Interestingly, they have come to quite different conclusions concerning the effects of private versus public schools on education. Rouse concludes that there are clear benefits to private schools in this sample whereas Witte does not. These differences appear to stem from different choices concerning the appropriate control group for analysis. Rouse uses those students who applied but were not selected for the program whereas Witte uses statewide average student outcomes. In terms of differencing out characteristics of the control and treatment groups, Rouse's approach seems clearly correct.

While important with respect to the issue of school vouchers and public policy, there are several grounds for supposing that the Milwaukee evidence has limited implications in terms of adducing the importance of interactions. As both Rouse and Witte are aware, since the majority of participants in the program went to one of only three different schools, the generality of any of the results is questionable. Further, it is important to remember that interactions, as conventionally understood, may not explain the differences either here or for differences between public and Catholic schools, which are discussed below. Differences in disciplinary standards or teacher expectations could potentially explain differences with the public schools independent of any interactions effects such as peer group influences. An interesting question for future research is therefore the determination of whether observed school differences occur due to interactions between students or due to alternative educational and disciplinary standards.

### 4. Classroom tracking

A standard problem in school organization is whether students should be tracked, i.e. segregated by ability and/or achievement across classes. A number of classroom experiments have been conducted in which educational outcomes for

students tracked by initial measures of ability of achievement are compared to students who are randomly assigned to classrooms.

One such experiment occurred in Montreal and has been analyzed by Henderson, Mieszkowski, and Sauvageau (1978). This paper analyzes data from French speaking students in Montreal in which children who were segregated on the basis of IQ tests administered in kindergarten and students who were randomly assigned to classes were compared in terms of achievement in grades 1-3. Henderson, Mieszkowski, and Sauvageau found significant effects from this type of classroom tracking. Interestingly, while randomization raised overall average performance, there was a clear diminution of the performance of students with higher test scores under random assignments. Hence randomization involves both redistribution as well as an increase in average achievement scores.

Unsurprisingly, ability grouping has also been studied quite extensively by education researchers, and has been a source of considerable controversy within the education literature. Slavin (1990) reviews a large number of tracking versus random assignment experiments in high schools and concludes that for secondary students "...between-class ability grouping plans have little or no effect on...achievement...at least as measured by standardized tests." (pg. 494) However, even this survey conclusion has been disputed by other education scholars as evidenced in the commentaries on that article. Our limited survey of the education literature suggests that there is little decisive evidence on this question, and that many of the studies are plagued by poor controls for individual characteristics; further, much of this literature seems laden with political concerns on the parts of researchers which make the assessment of the statistical analysis problematic. These techniques may also facilitate the determination of which neighborhood characteristics are relevant in generating interaction effects. Weisberg, Reagan, and Yankow (1999) show that an analysis employing a broad range of possible neighborhood controls can lead one to reject peer group and role model effects in favor of broader socioeconomic characteristics as the determinant of neighborhood effects.

5. Siblings

Matching comparisons have also been employed to directly control for unobserved family effects. Aaronson (1997,1998) proposes the use of sibling data to difference out unobserved family characteristics. This is possible under the assumption that the unobservable characteristics are constant within a family across time. He then identifies sibling pairs from the National Longitudinal Survey of Youth in which one sibling was exposed to a different neighborhood than another. This allows him to estimate models of differences in sibling outcomes which include differences in neighborhood characteristics. This estimation strategy is therefore equivalent to the standard one in panel data studies of differencing out unobserved fixed effects. Plotnick and Hoffman (1996) apply the same idea to a sample of sisters from the Panel Study of Income Dynamics and consider both continuous and discrete outcomes. Exploiting Chamberlain (1980) in order to eliminate unobserved fixed effects for binary choices, they find little evidence of neighborhood effects with respect to either out of wedlock births or any post-secondary education. This study finds no evidence of neighborhood effects on a particular income measure.

*b.* instrumental variables

Rather than employ data sets where interaction effects can be identified through the comparison of otherwise equivalent treatment and control groups, there has been a parallel literature which has tried to use more conventional econometric methods to deal with unobserved correlates.

1. Neighborhood socioeconomic influences

Evans, Oates, and Schwab (1992) appears to be the first study of neighborhood influences which formally accounts for the endogeneity of neighborhood residence. The analysis is specifically concerned with identifying the role of neighborhood characteristics on the probability of teen pregnancy. Using a probit framework, this probability is assumed to depend on both a range of individual characteristics as well as a variable which is the logarithm of the percentage of other students in an individual's high school who are categorized as "disadvantaged" as defined under guidelines of the Elementary and Secondary Education Act. In probit regressions which treat this measure as exogenous, this measure of disadvantaged schoolmates is shown to statistically significantly increase the probability of a teen pregnancy.

In order to deal with the possibility that the neighborhood characteristic measure is correlated with an unobserved individual characteristic, as would occur if parental quality is negatively associated with the neighborhood characteristic, Evans, Oates, and Schwab propose four instrumental variables each of which is measured at the level of the metropolitan area in which the secondary student lives: 1) the unemployment rate, 2) median family income, 3) the poverty rate and 4) the percentage of adults who are college graduates. The implicit assumption in this analysis is that the metropolitan area of residence is exogenous for families, although location within a metropolitan area is a choice variable. Employing these instruments, the contextual effect found in the univariate analysis disappears both in terms of magnitude and in terms of statistical significance.

**2. Catholic versus public schools**

Starting with Coleman, Hoffer and Kilgore (1982), a number of authors have studied the reasons why student performance in Catholic schools is on average superior to that found in public equivalents. A critical issue in evaluating the implications of this fact is determining whether the differences are due to self-

selection with respect to school enrollment versus something about differences in the school environment per se.

Evans and Schwab (1995) and Neal (1997) attempt to deal with the effect of self-selection by identifying instrumental variables which correlate with Catholic school choice but do not correlate with unobservable individual characteristics which would lead to better school performance. Neal's analysis seems especially comprehensive. He proposes two instruments which plausibly correlate with the decision to attend a Catholic school but not with unobserved individual characteristics which would lead to superior academic performance regardless of which school was attended: 1) the fraction of Catholic in county of residence population, which should correlate with tuition costs since higher percentages lead to greater Church subsidies to schools, and 2) the number of Catholic secondary schools per square mile within county, which should correlate negatively with transportation costs. The idea is tuition and transportation costs are plausibly correlated with the determinants of Catholic school choice without being correlated with an unobserved student quality variable. Neal finds that there are substantial educational gains for urban minorities who attend Catholic schools, but not for suburban students or whites in general.

## 9. Summary and conclusions

This chapter illustrates both the progress which has been made in utilizing interactions to understand economic phenomena as well as the many areas in which further research is required. In our judgment, there currently exists a good understanding of static interactions-based models both in terms of theory and econometrics. However, the empirical literature, while containing many insightful approaches to uncovering interactions, has yet to exploit a full structural estimation approach. Such a step is particularly important if one wishes to identify the presence of multiple equilibria. Further, there exist a number of

areas in terms of theory and econometric methodology which have yet to be fully examined. Three examples come readily to mind. First, the analysis of dynamic interaction models with endogenous neighborhood formation and their panel data analogs is still in its infancy. This analysis, fortunately, should have useful antecedents in the urban economics literature such as Miyao (1978). Second, the theoretical models of interactions currently treat the sources of interactions as a black box. In understanding phenomena such as social norms or culture, this is clearly inadequate; see the interesting analysis of Emirbayer and Goodwin (1997) for a discussion of the importance of properly accounting for the microfoundations of norms and culture. Third, the econometric literature has almost exclusively concentrated on global interactions, and so the analysis of identification and estimation needs to be extended to alternative interaction structures. Therefore, we are very confident that interactions-based models will continue to prove to be a productive area of research for methodologists and empiricists alike.

## Mathematical Appendix

### 1. Properties of binary choices made under a social planner.

#### *i.* Basics

Recalling the discussion in section 2.*v* in the text we consider a population of $I$ individuals whose choices, as determined by a social planner, follow the probability model

$$\mu(\underset{\sim}{\omega}) = exp(\beta(\sum_i (u(\omega_i, \underset{\sim}{Z}_i) + S(\omega_i, \underset{\sim}{Z}_i, \underset{\sim}{\omega}_{-i})))/Z_I \tag{A.1}$$

In this case, $\underset{\sim}{\omega}_{-i}$ is substituted for $\mu_i^e(\underset{\sim}{\omega}_{-i})$ in the social utility terms of the noncooperative problems and $Z_I$ is a normalizing constant. For the case of symmetric global interactions ($J_{i,j} = \frac{J}{I}$), employing the same transformations as done in the noncooperative case means that this probability may be rewritten as

$$\mu(\underset{\sim}{\omega}) = exp(\beta(\sum_i h_i\omega_i + \frac{J}{2I}(\sum_i \omega_i)^2))/Z_I \tag{A.2}$$

where the normalizing constant $Z_I$ is

$$\sum_{\nu_1 \in \{-1,1\}} \cdots \sum_{\nu_I \in \{-1,1\}} exp(\beta(\sum_{i=1}^{I} h_i\nu_i + \frac{J}{2I}(\sum_{i=1}^{I} \nu_i)^2)). \tag{A.3}$$

This equation corresponds to eq. (45) in the text.

In order to analyze this model, we make use of the following identity

$$exp(a^2) = (2\pi)^{-1/2} \int_{-\infty}^{\infty} exp(-\frac{x^2}{2} + 2^{1/2}xa)dx. \tag{A.4}$$

This identity can be verified immediately by dividing both sides of the expression by $exp(a^2)$ and recalling that the integral of the probability density of a normal

$(2^{1/2}a,1)$ random variable over its support is 1. Using the change of variable $y = x(\frac{\beta J}{I})^{1/2}$, it must be the case that

$$\mu(\underset{\sim}{\omega}) = (\frac{I}{2\pi\beta J})^{1/2} \int_{-\infty}^{\infty} exp(-\frac{y^2 I}{2\beta J}) \prod_i exp((y+\beta h_i)\omega_i) dy / Z_I \tag{A.5}$$

where

$$Z_I = (\frac{I}{2\pi\beta J})^{1/2} \int_{-\infty}^{\infty} exp(-\frac{y^2 I}{2\beta J}) \prod_i M(y+\beta h_i) dy \tag{A.6}$$

and

$$M(s) = exp(s) + exp(-s) \tag{A.7}$$

Notice that

$$\int_{-\infty}^{\infty} exp(-\frac{y^2 I}{2\beta J}) \prod_i M(y+\beta h_i) dy = \int_{-\infty}^{\infty} exp(I H_I(y)) dy \tag{A.8}$$

where

$$H_I(y) = \frac{1}{I} \sum_i ln(exp(v_i(1)) + exp(v_i(-1))) \tag{A.9}$$

and

$$v_i(\omega_i) = \beta h_i \omega_i + y\omega_i - \frac{y^2}{2\beta J}. \tag{A.10}$$

Notice that if $h_i = h \ \forall \ i$, then $H_I(y)$ does not depend on $I$.

It is shown rigorously by Amaro de Matos and Perez (1991) that as $I \Rightarrow \infty$, integrals of the form

$$\int_{-\infty}^{\infty} exp(I H_I(y)) dy \tag{A.11}$$

"pack" all mass onto the global maximizing point

$$y^* = argmax_y(H(y)) \tag{A.12}$$

where

$$H(y) = \int_{-\infty}^{\infty} ln(exp(\beta h + y - \frac{y^2}{2\beta J}) + exp(-\beta h - y - \frac{y^2}{2\beta J}))dF(h) \tag{A.13}$$

and $F(h)$ is the cumulative distribution function of $h$. One therefore expects (and can prove) that

$$y_I^* = argmax_y(H_I(y)) \Rightarrow y^* = argmax_y(H(y)) \tag{A.14}$$

Simple algebra reveals that the first order condition for the maximum of $H(\beta Jm)$ over $m$ is

$$m = \int_{-\infty}^{\infty} tanh(\beta h + \beta Jm)dF(h) \tag{A.15}$$

When $h_i = h \ \forall \ i$, this equation also holds for the expected value of each individual $i$ and hence the sample average by symmetry, which gives us Theorem 4.

Finally, this result suggests that if we replace the integral over $y$ in eq. (A.5) with a Dirac delta function whose mass is at $y^*$, we can obtain an approximate probability for the system of the form

$$\mu(\underset{\sim}{\omega}) = (\frac{I}{2\pi\beta J})^{1/2} exp(-\frac{y^{*2}I}{2\beta J}) \prod_i exp((y^* + \beta h_i)\omega_i)/Z_I \tag{A.16}$$

where $Z_I$ is a normalizing constant.

*ii.* Asymptotic moments: Proof of Theorem 11.

Let $E(\omega_i)$ denote the expectation of $\omega_i$ with respect to the probability measure (A.2). In order to determine the behavior of sample averages as $I \Rightarrow \infty$, we again consider the case where $h_i = h \ \forall \ i$. Notice that the argument of the previous section implies that

$$lim_{I \Rightarrow \infty} E(\omega_i) = \frac{lim_{I \Rightarrow \infty} \int_{-\infty}^{\infty} exp(IH_1(y)) G_1(y) dy}{lim_{I \Rightarrow \infty} \int_{-\infty}^{\infty} exp(IH_1(y)) dy} \tag{A.17}$$

where

$$H_1(y) = ln(M(\beta h + y)) - \frac{y^2}{2\beta J} \tag{A.18}$$

and

$$G_1(y) = \frac{exp(\beta h + y) - exp(-\beta h - y)}{M(\beta h + y)} = \frac{M'(\beta h + y)}{M(\beta h + y)} \tag{A.19}$$

We employ LaPlace's method (Kac (1968) pg. 248 or Ellis (1985), pp. 38, 50-51) to obtain the limiting values of these integrals. As described above, intuitively, all mass in these integrals gets packed onto the global maximizer $y^*$.

We restate the following useful result which is proven in Murray (1984) pg. 34.

**Approximation Theorem.**

Let $H(t)$ be a function on the interval $(a,b)$ which takes a global maximum at a point $\alpha$ in the interval and let $H(t)$ be smooth enough to possess a second-order Taylor expansion at point $\alpha$ with $H''(\alpha) < 0$. Let $G(t)$ denote a continuous

function. Then

$$\int_{-\infty}^{\infty} G(t)exp(IH(t))dt =$$

$$exp(IH(\alpha))G(\alpha)(\frac{-2\pi}{IH''(\alpha)})^{1/2} + O(I^{-3/2}) \tag{A.20}$$

This formula states, in a precise way, the sense in which the mass of the integral piles up at the maximizer $\alpha$ as $I \Rightarrow \infty$. Using this formula, letting $\alpha = y$,

$$\frac{\int_{-\infty}^{\infty} exp(IH_1(y))G_1(y)dy}{\int_{-\infty}^{\infty} exp(IH_1(y))dy} =$$

$$\frac{exp(IH_1(\alpha))G_1(\alpha)(\frac{-2\pi}{IH_1''(\alpha)})^{1/2} + O(I^{-3/2})}{exp(IH_1(\alpha))(\frac{-2\pi}{IH_1''(\alpha)})^{1/2} + O(I^{-3/2})} \tag{A.21}$$

which is easily seen to converge to $G_1(\alpha)$ as $I \Rightarrow \infty$. Hence we have

$$lim_{I \Rightarrow \infty} E(\omega_i) = G_1(y^*) = m^* = \frac{exp(\beta h + y^*) - exp(-\beta h - y^*)}{M(\beta h + y^*)} =$$

$$\frac{M'(\beta h + y^*)}{M(\beta h + y^*)} = tanh(\beta h + y^*) = tanh(\beta h + \beta J m^*) \tag{A.22}$$

where $y^* = \beta J m^*$.

The problem that $m^*$ solves appears mysterious at first glance. However, there is an interesting connection between our solution to the behavior of a social planner and the maximization of social surplus as analyzed in McFadden (1981). Following McFadden, social surplus will equal $\sum_i (u(\omega_i, \underset{\sim}{X}_i) - \frac{J}{2}(\omega_i - \bar{\omega}_I)^2)$. If all agents have common characteristics $\underset{\sim}{X}_i$, then following (A.2), the probability of

the social surplus can be expressed as a function of $G(\underset{\sim}{\omega}) = \sum_i h\omega_i + \frac{J}{2I}(\sum_i \omega_i)^2$. Then it can be shown (Brock (1993)) that

$$\beta(lim_{I\Rightarrow\infty} E(max_{\underset{\sim}{\omega}} I^{-1} G(\underset{\sim}{\omega}))) = lim_{I\Rightarrow\infty}(I^{-1} lnZ_I) =$$

$$max_y \, ln(exp(-\frac{y^2}{2\beta J}) M(\beta h + y)) =$$

$$max_m \, ln(exp(-\frac{(\beta J m)^2}{2\beta J}) M(\beta h + \beta J m)) \tag{A.23}$$

As would be expected, one maximizes a notion of social welfare in the large economy limit in order to find the socially optimal states.

Now that the expected value for each choice has been analyzed, we can consider laws of large numbers for data generated in this environment. First, we consider the sample mean, $\overline{\omega}_I = I^{-1}\sum_i \omega_i$. Notice that the limiting behavior of the sample mean in distribution ($\Rightarrow_d$) can be inferred from weak convergence ($\Rightarrow_w$) since weak convergence necessarily implies convergence in distribution (see Lukacs (1975) pg. 9 for a typical proof.) By Tchebychev's inequality,

$$\mu(\,|\,\overline{\omega}_I - m^*\,|\,\geq \epsilon) \;\leq \frac{Var(\overline{\omega}_I - m^*)}{\epsilon^2} \tag{A.24}$$

so it is sufficient to prove $lim_{I\Rightarrow\infty} Var(\overline{\omega}_I - m^*) = 0$. To do this, it is sufficient to show that $I^{-2}\sum_i \omega_i \sum_j \omega_j \Rightarrow_w m^{*2}$. However, this can be verified (after considerable algebra) by computing $I^{-2}\sum_i \omega_i \sum_j \omega_j$ directly and using LaPlace's method as employed in Murray above to verify that $I^{-2}\sum_i \omega_i \sum_j \omega_j \Rightarrow_w tanh(\beta h + y^*)^2 = m^{*2}$. This proves Theorem 12.

### *iii.* Maximum likelihood theory

Consider $g = 1...G$ distinct neighborhoods with observations $\underset{\sim}{X}_{i,g}$, $i = 1...I$

and $\overline{\omega}_g = I^{-1}\sum_i \omega_{i,g}$ available for each $g$. Define the likelihood function for the data from these neighborhoods as $\prod_g \mu(\underset{\sim}{\omega}_g)$ where $\underset{\sim}{\omega}_g = (\omega_{1,g}...\omega_{I,g})$. When choices are consistent with the solution to a social planners problem, the likelihood function within each neighborhood will have the form

$$\mu(\underset{\sim}{\omega}_g) \sim exp(\sum_i (\tfrac{1}{2}c'\underset{\sim}{X}_{i,g} + \tfrac{J}{I}(\sum_j \omega_{j,g}))\omega_{i,g}) \tag{A.25}$$

which can be rewritten as

$$(\frac{I}{2\pi\beta J})^{1/2}\int_{-\infty}^{\infty} exp(-\overline{\omega}_g^2\frac{IJ}{2})\prod_i exp((\tfrac{1}{2}c'\underset{\sim}{X}_{i,g} + J\overline{\omega}_g)\omega_{i,g})d\overline{\omega}_g \tag{A.26}$$

Define the parameter vector $\theta = (c,J)$. One can consider the mean log likelihood over all observations

$$\frac{1}{GI}\sum_g ln\mu(\underset{\sim}{\omega}_g). \tag{A.27}$$

For large $I$, and letting $F(x) = \dfrac{exp(x)}{1+exp(x)}$ the density of this likelihood will approximately equal

$$\begin{aligned}\frac{1}{G}\sum_g \frac{1}{I}\sum_i ((\frac{1+\omega_{i,g}}{2})ln(F(c'\underset{\sim}{X}_{i,g} + 2J\mu_{I,g})) + \\ (\frac{1-\omega_{i,g}}{2})ln(F(-(c'\underset{\sim}{X}_{i,g} + 2J\mu_{I,g}))))\end{aligned} \tag{A.28}$$

where

$$\mu_{I,g} = argmax(H_{I,g}(\mu)) \tag{A.29}$$

and

$$H_{I,g}(\mu) =$$

$$-\frac{\mu^2 J}{2} + \frac{1}{I}\left(\sum_i ln(exp(\tfrac{1}{2}c'\underset{\sim}{X}_{i,g} + J\mu)) + exp(-\tfrac{1}{2}c'\underset{\sim}{X}_{i,g} - J\mu)\right) \qquad (A.30)$$

Note that $H_{I,g}$ converges to

$$H_g(\mu) =$$

$$-\frac{\mu^2 J}{2} + \int ln(exp(\tfrac{1}{2}c'\underset{\sim}{X}_{i,g} + J\mu) + exp(-\tfrac{1}{2}c'\underset{\sim}{X}_{i,g} - J\mu))dF_g(\underset{\sim}{X}_{i,g}) \quad (A.31)$$

so that under regularity conditions such as those described in Newey and McFadden (1994) pg. 2121 it must be the case that

$$\mu_{I,g} \Rightarrow_w \mu_g = argmax(H_G(\mu)) \qquad (A.32)$$

Notice that the naive estimator introduced in Section 3 inserts $\overline{\omega}_g$ in place of $\mu_{I,g}$ in the sample likelihood (A.28) above and selects $\theta$ to maximize the modified sample log likelihood function. This means that the naive estimator does not allow the data to directly address the possibility of discontinuous neighborhood responses because the standard maximum likelihood theory in logistic models yields a strictly concave optimization problem. Hence the optimization problem will be continuous in parameters such as the distribution function of individual characteristics, $F_g(\underset{\sim}{X}_{i,g})$.

This suggests that one might wish to modify this log likelihood by adding a penalty function of the form

$$\frac{A}{G}\sum_g (\overline{\omega}_g - \mu_{I,g})^2 \qquad (A.33)$$

$A = 0$ will correspond to the naive estimator. Intuitively, as $A$ increases the penalty will push the parameter estimates towards those of the complete estimator, i.e. one which accounts for the relationship between the neighborhood

characteristics and neighborhood mean behavior.

**2. Proof of Theorem 5.**

For a given parameter set $(k,c,d,J)$, assume by way of contradiction that there exists an alternative $(\overline{k}\,,\overline{c}\,,\overline{d}\,,\overline{J}\,)$ such that on $supp(\underset{\sim}{X},\underset{\sim}{Y},m^e)$ we have

$$(k-\overline{k})+(c'-\overline{c}')\underset{\sim}{X}_i+(d'-\overline{d}')\underset{\sim}{Y}_{n(i)}+(J-\overline{J}\,)m^e_{n(i)}=0 \qquad \text{(A.34)}$$

and

$$m^e_{n(i)}=m_{n(i)}=\int \omega_i dF(\omega_i \mid k+c'\underset{\sim}{X}+d'\underset{\sim}{Y}_{n(i)}+Jm_{n(i)})dF_{\underset{\sim}{X}\mid \underset{\sim}{Y}_{n(i)}}=$$

$$\int \omega_i dF(\omega_i \mid \overline{k}\,+\overline{c}'\underset{\sim}{X}+\overline{d}'\underset{\sim}{Y}_{n(i)}+\overline{J}m_{n(i)})dF_{\underset{\sim}{X}\mid \underset{\sim}{Y}_{n(i)}}. \qquad \text{(A.35)}$$

Notice the Proposition is true if it is the case that $J-\overline{J}$ is zero. Otherwise, $\underset{\sim}{X}_i$ and $\underset{\sim}{Y}_{n(i)}$ would lie in a proper linear subspace of $R^{r+s}$ which violates Assumption *i*. Equation (A.34) implies that for elements of $supp(\underset{\sim}{X},\underset{\sim}{Y},m^e)$, conditional on $\underset{\sim}{Y}_{n(i)}$

$$(c'-\overline{c}')\underset{\sim}{X}_i=\rho(\underset{\sim}{Y}_{n(i)}) \qquad \text{(A.36)}$$

where $\rho(\underset{\sim}{Y}_{n(i)})=-(k-\overline{k})-(d'-\overline{d}')\underset{\sim}{Y}_{n(i)}-(J-\overline{J}\,)m^e_{n(i)}$. (A.36) must hold for all neighborhoods, including $n_0$ as described in Assumption *iv.* of the Theorem. This would mean that, conditional on $\underset{\sim}{Y}_{n_0}$, and given that $\underset{\sim}{X}_i$ cannot contain a constant by Assumption *iii.*, that $\underset{\sim}{X}_i$ is contained in a proper linear subspace of $R^r$ and therefore violates the Assumption *iv.* of the Proposition. Hence, $c$ is identified.

Given identification of $c$, (A.34) now implies, if $J\neq\overline{J}$, that $m^e_{n(i)}$ is a linear function of $\underset{\sim}{Y}_{n(i)}$, unless $(d'-\overline{d}')$ and/or $m^e_{n(i)}$ is always equal to zero. The

latter is ruled out by Assumption *vi*. Linear dependence of $m^e_{n(i)}$ on $\underset{\sim}{Y}_{n(i)}$ when $(d' - \overline{d}') \neq 0$ contradicts the combination of the requirement that support of $m^e_{n(i)}$ is $[-1,1]$ with Assumption *v*., that the support of each component of $\underset{\sim}{Y}_{n(i)}$ is unbounded, since $\underset{\sim}{Y}_{n(i)}$ can, if it is unbounded, assume values with positive probability that violate the bounds on $m^e_{n(i)}$. So, $J$ is identified. If $J$ is identified and $(d' - \overline{d}') \neq 0$, then (A.34) requires that

$$(d' - \overline{d}')\underset{\sim}{Y}_{n(i)} = -(k - \overline{k}) \tag{A.37}$$

for all $\underset{\sim}{Y}_{n(i)} \in supp(\underset{\sim}{Y}_{n(i)})$. This implies, since by Assumption *iii*. $\underset{\sim}{Y}_{n(i)}$ does not contain a constant, that $supp(\underset{\sim}{Y}_{n(i)})$ is contained in a proper linear subspace of $R^s$, which contradicts condition *ii*. of the Theorem. Therefore, $d' = \overline{d}'$. This immediately implies that $k = \overline{k}$ and the Theorem is verified.

**3. Proof of Theorem 7.**

As is done in the text, $A$ denotes the parameter set $(k,c,d,J)$ and the conditional mean function is $H = k + c'\underset{\sim}{X}_i + d'\overline{\underset{\sim}{X}}_{n(i)} + JG(m)$. To verify the theorem, it is necessary to show that the components of the gradient vector

$$d_A H = \frac{\partial H}{\partial A} + \frac{\partial H}{\partial m}\frac{\partial m}{\partial A} \tag{A.38}$$

define a linearly independent collection of functions of $\underset{\sim}{X}_i$ and $\overline{\underset{\sim}{X}}_{n(i)}$ on $supp(\underset{\sim}{X}_i, \overline{\underset{\sim}{X}}_{n(i)})$. Differentiation implies the following, which we will use.

$$\frac{\partial H}{\partial A} = (1, \underset{\sim}{X}_i, \overline{\underset{\sim}{X}}_{n(i)} G(m)) \tag{A.39}$$

$$\frac{\partial H}{\partial m} = J(1 + \xi\frac{dg(m)}{m}) \tag{A.40}$$

Since $J \neq 1$ and $g$ is $C^2$, the neighborhood $N_\epsilon$ can always be chosen so that an implicit function $m(\underset{\sim}{\bar{X}}_{n(i)},A,\xi)$ exists. Also, define the function $J(m,\xi) = J(1+\xi\frac{dg(m)}{m})$.

Rewrite the gradient as

$$d_A H = \frac{1}{1-J(m,\xi)}(1,\underset{\sim}{X}_i+J(m,\xi)(\underset{\sim}{\bar{X}}_{n(i)}-\underset{\sim}{X}_i),\underset{\sim}{\bar{X}}_{n(i)},m+\xi g(m)) \quad \text{(A.41)}$$

If $\xi$ is close enough to zero, $J(m,\xi)$ cannot equal 1 since $J \neq 1$ by Assumption *iii.* This is a vector proportional to the form $v = (1,v_2(\underset{\sim}{X}_i,\underset{\sim}{\bar{X}}_{n(i)}),v_3(\underset{\sim}{\bar{X}}_{n(i)}),v_4(\underset{\sim}{\bar{X}}_{n(i)}))$. Notice that we have eliminated $m$ since its implicit function solution makes it a function of $\underset{\sim}{\bar{X}}_{n(i)}$. In order to show linear independence, we must verify that

$$a_1 + a_2 v_2(\underset{\sim}{X}_i,\underset{\sim}{\bar{X}}_{n(i)}) + a_3 v_3(\underset{\sim}{\bar{X}}_{n(i)}) + a_4 v_4(\underset{\sim}{\bar{X}}_{n(i)}) = 0 \quad \text{(A.42)}$$

implies that $a_1 = a_2 = a_3 = a_4 = 0$.

Since only $v_2$ depends on $\underset{\sim}{X}_i$, (A.42) can only hold if $a_2 = 0$; otherwise Assumption *ii.* would be violated. Further, if $a_4 = 0$, then Assumption *i.* is violated. This is true because $v_3(\underset{\sim}{\bar{X}}_{n(i)})$ is proportional to $\underset{\sim}{\bar{X}}_{n(i)}$. We can therefore, without loss of generality assume $a_4 = -1$.

The condition for linear independence can now be written as

$$m(\underset{\sim}{\bar{X}}_{n(i)},A,\xi) + \xi g(m(\underset{\sim}{\bar{X}}_{n(i)},A,\xi)) = a_1 + a_3\underset{\sim}{\bar{X}}_{n(i)} \quad \text{(A.43)}$$

We pair this with the self-consistency condition written as

$$m(\underset{\sim}{\bar{X}}_{n(i)},A,\xi) =$$

$$k + (c'+d')\underset{\sim}{\bar{X}}_{n(i)} + J(m(\underset{\sim}{\bar{X}}_{n(i)},A,\xi)) + \xi g(m(\underset{\sim}{\bar{X}}_{n(i)},A,\xi)). \quad \text{(A.44)}$$

We will verify that (A.43) and (A.44) lead to a contradiction when $\frac{dg}{dm}$ differs

across any two $m$ values, say $m_1$ and $m_2$. Since at least two such values must exist by Assumption *iv.*, this will complete the proof.

On the open set $O$ described by Assumption *iv.*, we can differentiate both these equations with respect to $\bar{\underset{\sim}{X}}_{n(i)}$, obtaining

$$(1+\xi\frac{dg(m(\bar{\underset{\sim}{X}}_{n(i)},A,\xi))}{dm})\frac{dm(\bar{\underset{\sim}{X}}_{n(i)},A,\xi)}{d\bar{\underset{\sim}{X}}_{n(i)}}=a_3 \tag{A.45}$$

and

$$\frac{dm(\bar{\underset{\sim}{X}}_{n(i)},A,\xi)}{d\bar{\underset{\sim}{X}}_{n(i)}}(1-J(1+\xi\frac{dg(m(\bar{\underset{\sim}{X}}_{n(i)},A,\xi))}{dm}))=(c'+d') \tag{A.46}$$

Equating $\frac{dm(\bar{\underset{\sim}{X}}_{n(i)},A,\xi)}{d\bar{\underset{\sim}{X}}_{n(i)}}$ across these expressions yields

$$(1-J(1+\xi\frac{dg(m(\bar{\underset{\sim}{X}}_{n(i)},A,\xi))}{dm}))a_3-(c+d)(1+\xi\frac{dg(m(\bar{\underset{\sim}{X}}_{n(i)},A,\xi))}{dm})=0 \tag{A.47}$$

or

$$\xi\frac{dg(m(\bar{\underset{\sim}{X}}_{n(i)},A,\xi)}{dm}((c+d)+Ja_3)+((c+d)+Ja_3-a_3)=0 \tag{A.48}$$

Recall that $\xi$ and $\frac{dg}{dm}$ are scalars, whereas $c$, $d$, and $a_3$ are $r\times 1$ vectors. By construction of $g$, we have the existence of two values of $m$, call them $m_1$ and $m_2$ such that in the population data, $\frac{dg}{dm}$ differs across them. Applying this component by component to (A.47), one can show that this implies that $(c+d)+Ja_3=0$. By (A.48), this means that $a_3=0$. But from (A.49), this would imply that

$$m(\bar{\underset{\sim}{X}}_{n(i)},A,\xi)+\xi g(m(\bar{\underset{\sim}{X}}_{n(i)},A,\xi))=a_1. \tag{A.49}$$

But this would contradict the part of Assumption *iv.* that $m_{n(i)}$ in the data is nonconstant. Therefore, the model and assumptions described by the Theorem require that the components of the gradient (A.38) are linearly independent when $\xi \neq 0$. Notice that when $\xi = 0$, the gradient will not be of full rank, because $m(\underset{\sim}{\bar{X}}_{n(i)}, A, 0)$ is linear in $\underset{\sim}{\bar{X}}_{n(i)}$. Hence the local nonidentification of the linear-in-means model can be perturbed away by a $C^2-$small change from $Jm$ to $Jm + \xi g(m)$, which completes the proof.

## Bibliography


Aaronson, D., (1998), "Using Sibling Data to Estimate the Impact of Neighborhoods on Children's Educational Outcomes," Journal of Human Resources, XXXIII, 915-946.

Aaronson, D., (1997), "Sibling Estimates of Neighborhood Effects," in Neighborhood Poverty: Policy Implications for Studying Neighborhoods, v. II, J. Brooks-Dunn, G. Duncan, and L. Aber, eds., New York: Russell Sage Foundation.

Ahn, H. and J. Powell, (1993), "Semiparametric Estimation of Censored Regression Models with a Nonparametric Selection Mechanism," Journal of Econometrics

Akerlof, G., (1997), "Social Distance and Social Decisions," Econometrica, 65, 1005-1028.

Akerlof, G. and R. Kranton, (1998), "Identity and Economics," mimeo, University of California at Berkeley.

Akerlof, G. and J. Yellen, (1994), "Gang Behavior, Law Enforcement, and Community Values," in Values and Public Policy, H. Aaron, T. Mann, and T. Taylor eds., Washington D.C.: Brookings Institution Press.

Akerlof, G., J. Yellen, and M. Katz, (1996), "An Analysis of Out-of-Wedlock Childbearing in the United States," Quarterly Journal of Economics, CXI, 277-318.

Alessie, R. and A. Kapteyn, (1991), "Habit Formation, Interdependent Preferences and Demographic Effects in the Almost Ideal Demand System," Economic Journal, 101, 404-419.

Amaro de Matos, J. M. G., and J. F. Perez, (1991), "Fluctuations in the Curie-Weiss Version of the Random Field Ising Model," Journal of Statistical Physics, 62, 3/4, 587-608.

Amemiya, T., (1985), Advanced Econometrics, Cambridge: Harvard University Press.

An, M. and N. Kiefer, (1995), "Local Externalities and Societal Adoption of Technologies," Journal of Evolutionary Economics, 5, 103-117.

Anderson, E., (1990), Streetwise, Chicago: University of Chicago Press.

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

Andreoni, J. and J. K. Scholz, (1998), "An Econometric Analysis of Charitable Giving with Interdependent Preferences," Economic Inquiry, XXXVI, 410-428.

Arthur, W. B., (1987), "Urban Systems and Historical Path Dependence," in Urban Systems and Infrastructure, R. Herman and J. Ausubel, eds., Washington D.C.: National Academy of Sciences/National Academy of Engineering.

Arthur, W. B., (1989), "Increasing Returns, Competing Technologies and Lock-In by Historical Small Events: The Dynamics of Allocation Under Increasing Returns to Scale," Economic Journal, 99, 116-131.

Ash, R., (1972), Real Analysis and Probability, Orlando: Academic Press.

Averintsev, M., (1970), "On a Method of Describing Discrete Parameter Fields," Problemy Peredachi Informatsii, 6, 100-109.

Axtell, R., R. Axelrod, J. Epstein, and M. Cohen, (1996), "Aligning Simulation Models: A Case Study and Results," Computational and Mathematical Organization Theory, 1:2, 123-141.

Banerjee, A., (1992), "A Simple Model of Herd Behavior," Quarterly Journal of Economics, CVII, 797-818.

Bauman, K., and L. Fisher, (1986), "On the Measurement of Friend Behavior in Research on Friend Influence and Selection: Findings from Longitudinal Studies of Adolescent Smoking and Drinking," Journal of Youth and Adolescence, 15, 345-353.

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

Becker, G., (1996), Accounting for Tastes, Cambridge: Harvard University Press.

Bell, A., (1995), "Dynamically Interdependent Preferences in a General Equilibrium Environment," mimeo, Department of Economics, Vanderbilt University.

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

Bénabou, R., (1996a), "Equity and Efficiency in Human Capital Investment:

The Local Connection," Review of Economic Studies, 62, 237-264.

Bénabou, R., (1996b), "Heterogeneity, Stratification, and Growth: Macroeconomic Effects of Community Structure," American Economic Review 86, 584-609.

Bernheim, D., (1994), "A Theory of Conformity," Journal of Political Economy, 5, 102, 841-877.

Bertrand, M., E. Luttmer, and S. Mullainathan, (1998), "Network Effects and Welfare Cultures," working paper, Harvard University.

Billingsley, P., (1968), Convergence of Probability Measures, New York: John Wiley and Sons.

Binder, M. and M. H. Pesaran, (1998a), "Decision Making in the Presence of Heterogeneous Information and Social Interactions," working paper, University of Maryland.

Bikhchandani, S., D. Hirshleifer, and I. Welch, (1992), "A Theory of Fads, Fashion, Custom, and Cultural Exchange as Information Cascades," Journal of Political Economy, 100, 992-1026.

Binder, M. and M. H. Pesaran, (1998b), "Life-Cycle Consumption Under Social Interactions," working paper, University of Maryland.

Blalock, H., (1984), "Contextual-Effects Models: Theoretical and Methodological Issues," Annual Review of Sociology, 10, 353-372.

Blau, D. and O. Duncan, (1967), The American Occupational Structure, New York: John Wiley and Sons.

Blume, L., (1993), "The Statistical Mechanics of Strategic Interaction," Games and Economic Behavior, 5, 387-424.

Blume, L, (1995), "The Statistical Mechanics of Best-Response Strategy Revision," Games and Economic Behavior, 11, 111-145.

Blume, L., (1997), "Population Games," in The Economy as a Complex Evolving System II, W. B. Arthur, S. Durlauf and D. Lane, eds. Menlo Park: Addison-Wesley.

Blume, L. and S. Durlauf, (1998a), "The Interactions-Based Approach to Socioeconomic Behavior," mimeo, University of Wisconsin.

Blume, L. and S. Durlauf, (1998b), "Equilibrium Concepts for Social Interaction Models," mimeo, Cornell University.

Borjas, G., (1992), "Ethnic Capital and Intergenerational Income Mobility," Quarterly Journal of Economics, CVII, 123-150.

Borjas, G., (1995), "Ethnicity, Neighborhoods, and Human-Capital Externalities," American Economic Review, 85, 365-390.

Bretagnolle, J. and C. Huber-Carol, (1988), "Effects of Omitting Covariates in Cox's Model for Survival Data," Scandinavian Journal of Statistics, 15, 125-138.

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

Brewster, K., (1994b), "Neighborhood Context and the Transition to Sexual Activity Among Black Women," Demography, 31, 603-614.

Brock, W., (1993), "Pathways to Randomness in the Economy: Emergent Nonlinearity and Chaos in Economics and Finance," Estudios Economicos, 8, 1, 3-55 and Social Systems Research Institute Reprint #410, Department of Economics, University of Wisconsin at Madison. Reprinted in Dechert, W., ed., 1996, Chaos Theory in Economics: Methods, Models and Evidence, Edward Elgar Cheltenham UK.

Brock, W., and C. Hommes, (1998), "Rational Routes to Randomness," Econometrica, 65, 1059-1096.

Brock, W., and S. Durlauf, (1995), "Discrete Choice with Social Interactions," mimeo, University of Wisconsin at Madison.

Brock, W., and S. Durlauf, (1998b), "A Formal Model of Theory Choice in Science," Economic Theory, forthcoming.

Brooks-Gunn, J., G. Duncan, P. Klebanov, and N. Sealand, (1993), "Do Neighborhoods Affect Child and Adolescent Development?," American Journal of Sociology, 99, 353-395.

Brown, B., (1990), "Peer Groups and Peer Cultures," in At the Threshold, S. Feldman and G. Elliott eds., Cambridge: Harvard University Press.

Brown, B., D. Clasen, and S. Eicher, (1986), "Perceptions of Peer Pressure, Peer Conformity Dispositions, and Self-Reported Behavior Among Adolescents," Developmental Psychology, 22, 521-530.

Bryant, J. (1983), "A Simple Rational Expectations Keynes-Type Model," Quarterly Journal of Economics, XCVIII, 525-529.

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, 262-333.

Caplin, A. and J. Leahy, (1994), "Business as Usual, Market Crashes, and Wisdom After the Fact," American Economic Review 84, 548-565.

Case, A., (1992), "Neighborhood Influence and Technological Change," Regional Science and Urban Economics, 22, 491-508.

Case, A., and L. Katz, (1991), "The Company You Keep: The Effects of Family and Neighborhood on Disadvantaged Families," National Bureau of Economic Research Working Paper no. 3705.

Casella, A. and J. Rauch, (1997), "Anonymous Market and Group Ties in International Trade," NBER Working Paper no. 6186.

Casella, A. and J. Rauch, (1998), "Overcoming Informational Barriers to International Resource Allocation: Prices and Group Ties" NBER Working Paper no. 6628.

Chamberlain, G., (1984), "Panel Data," in Z. Griliches and M. Intrilligator, eds., Handbook of Econometrics: Volume II, Amsterdam: North-Holland.

Clark, A. and A. Oswald, (1996), "Satisfaction and Comparison Income," Journal of Public Economics, 61, 359-381.

Clark, A. and A. Oswald, (1998), "Comparison-Concave Utility and Following Behaviour in Social and Economic Settings," Journal of Public Economics, 70, 133-155.

Cole, H., G. Mailath, and A. Postlewaite, (1992), "Social Norms, Savings Behavior, and Growth," Journal of Political Economy, 100, 1092-1125.

Coleman, J., (1988), "Social Capital in the Creation of Human Capital," American Journal of Sociology, 94 (supplement), S95-S120.

Coleman, J., (1990), Foundations of Social Theory, Cambridge: Harvard University Press.

Coleman, J., E. Campbell, J. Hobson, J. McPartland, A. Mood, F. Weinfeld, and

R. York, (1966), Equality of Educational Opportunity, Washington D.C.: US Government Printing Office.

Coleman, J., T. Hoffer and S. Kilgore, (1982), High School Achievement: Public, Private and Catholic Schools Compared, New York: Basic Books.

Conley, T. and G. Topa, (1999), "Socio-Economic Distance and Spatial Patterns in Unemployment, mimeo, New York University.

Conlisk, J., (1976), "Interactive Markov Chains," Journal of Mathematical Sociology, 4, 157-185.

Cooper, R., and A. John, (1988), "Coordinating Coordination Failures in Keynesian Models," Quarterly Journal of Economics, CIII, 441-464.

Cooper, S., (1998), "A Positive Theory of Income Redistribution," Journal of Economic Growth, 3, 171-195.

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

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

Crane, J., (1991b), "Effects of Neighborhoods on Dropping out of School and Teenage Childbearing," in The Urban Underclass, C. Jencks and P. Peterson eds., Washington DC: Brookings Institution Press.

Cross, P. and C. Manski, (1999), "Regressions, Long and Short," mimeo, Northwestern University.

Cutler, D. and E. Glaeser, (1997), "Are Ghettos Good or Bad?," Quarterly Journal of Economics, CXII, 827-872.

Darrough, M., R. Pollak, and T. Wales, (1983), "Dynamic and Stochastic Structure: An Analysis of Three Time Series of Household Budget Studies," Review of Economics and Statistics, 65, 274-281.

Dasgupta, P., (1995), "The Population Problem: Theory and Evidence," Journal of Economic Literature, 33, 1879-1982.

Dasgupta, P. and P. David, (1994), "Towards a New Economics of Science," Research Policy, 23, 487-521.

David, P. (1985), "Clio and the Economics of QWERTY," American Economic Review, 75, 332-337.

David, P., (1998), "The Collective Cognitive Performance of "Invisible Colleges," working paper, Oxford University.

Debreu, G., (1952), "A Social Equilibrium Existence Theorem," Proceedings of the National Academy of Sciences, 38, 886-893.

Duesenberry, J. (1949), Income, Saving, and the Theory of Consumer Behavior, Cambridge: Harvard University Press.

Duncan, G. and S. Raudenbusch, (1998), "Neighborhoods and Adolescent Development: How Can We Assess the Links?," working paper, Northwestern University.

Duneier, M. and H. Molotch, (1999), "Talking City Trouble: Interactional Vandalism, Social Inequality, and the "Urban Interaction Problem"," mimeo, University of Wisconsin and forthcoming, American Sociological Review.

Durlauf, S., (1993), "Nonergodic Economic Growth," Review of Economic Studies, 60, 349-366.

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: Proceedings of the Ninth International Symposium on Economic Theory and Econometrics, W. Barnett, G. Gandolfo, and C. Hillinger, eds., Cambridge University Press.

Durlauf, S., (1997), "Statistical Mechanics Approaches to Socioeconomic Behavior," in The Economy as a Complex Evolving System II, W. B. Arthur, S. Durlauf and D. Lane, eds. Redwood City: Addison-Wesley.

Durlauf, S. and D. Quah, (1999), "The New Empirics of Economic Growth," with D. Quah, in Handbook of Macroeconomics, J. Taylor and M. Woodford eds., North-Holland, forthcoming.

Durlauf, S. and J. Walker, (1998), "Social Interactions and Fertility Transitions," mimeo, University of Wisconsin at Madison.

Elbers, C. and G. Ridder, (1982), "True and Spurious Duration Dependence: The Identifiability of the Proportional Hazards Model," Review of Economic Studies,

49, 403-410.

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

Emirbayer, M. and J. Goodwin, (1997), "Network Analysis, Culture, and the Problem of Agency," American Journal of Sociology, 99, 1411-1454.

Epple, D. and T. Romer, (1991), "Mobility and Redistribution," Journal of Political Economy, 99, 828-858.

Epple, D. and H. Sieg, (1999), "Estimating Equilibrium Models of Local Jurisdictions," mimeo, Carnegie-Mellon University and forthcoming, Journal of Political Economy.

Epstein, J. and R. Axtell, (1996), Growing Artificial Societies: Social Science from the Bottom Up, Cambridge: MIT Press.

Erdélyi, A., (1956), Asymptotic Expansions, New York: Dover Publications.

Evans, W. and R. Schwab, (1995), "Finishing High School and Starting College: Do Catholic Schools Make A Difference?," Quarterly Journal of Economics, CX, 909-940.

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

Farrell, J. and G. Saloner, (1985), "Standardization, Compatibility and Innovation," Rand Journal of Economics, 16, 70-83.

Fernandez, R. and R. Rogerson, (1996), "Income Distribution, Communities, and the Quality of Public Education," Quarterly Journal of Economics, 111, 135-164.

Fingleton, B., (1999), "Economic Geography with Spatial Econometrics: A "Third Way" to Analyze Economic Development and "Equilibrium" with Application to the EU Regions," Working Paper no. 99/21, Department of Economics, European University Institute.

Föllmer, H., (1974), "Random Economies with Many Interacting Agents," Journal of Mathematical Economics, 1, 51-62.

Frank, R., (1985), Choosing the Right Pond, New York: Oxford University Press.

Freedman, D., (1987), "As Others See Us: A Case Study in Path Analysis," Journal of Educational Statistics, 12, 101-223.

Freedman, D., (1991), "Statistical Models and Shoe Leather," in Sociological Methodology, P. Marsden, ed., Cambridge: Basil Blackwell.

Freedman, D., S. Klein, P. Sacks, J. Smythe, C. Everett, "Ecological Inference and Voting Rights (with discussion)," Evaluation Review, 15, 673-711.

Freedman, D., S. Klein, M. Ostland, and M. Roberts, (1998), "On "Solutions" to the Ecological Inference Problem," Technical Report, Department of Statistics, UC Berkeley.

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

Glaeser, E. and J. Scheinkman, (1998), "Measuring Social Interactions," mimeo, Department of Economics, Harvard University.

Goering, J., (1996), "Expanding Housing Choices for HUD-Assisted Families: First Biennial Report on The Moving to Opportunity for Fair Housing Demonstration," Department of Housing and Urban Development, Office of Policy Development and Research.

Goodman, L., (1953), "Ecological Regression and the Behavior of Individuals," American Sociological Review, 18, 663-666.

Goolsbee, A. and P. Klenow, (1998). "Evidence on Network and Learning Externalities in the Diffusion of Home Computers," mimeo, Graduate School of Business, University of Chicago.

Gottfredson, M. and T. Hirschi, (1990), A General Theory of Crime, Stanford: Stanford University Press.

Granovetter, M. and R. Soong, (1986), "Threshold Models of Interpersonal Effects in Consumer Demand," Journal of Economic Behavior and Organization, 7, 83-99.

Granovetter, M. and R. Soong, (1988), "Threshold Models of Diversity: Chinese Restaurants, Residential Segregation, and the Spiral of Silence," in Sociological Methodology, vol. 18, C. Clogg, ed., Cambridge: Basil Blackwell.

Granovetter, M., (1995), Getting a Job, Second Edition, Chicago: University of Chicago Press.

Hansen, L. and T. Sargent, (1991), Rational Expectations Econometrics, Boulder: Westview Press.

Hauser, R., (1970), "Context and Consex: A Cautionary Tale," American Journal of Sociology, 75, 645-664.

Haveman, R. and B. Wolfe, (1994), Succeeding Generations, New York: Russell Sage Foundation.

Heckman J., (1978), "Simple Statistical Models for Discrete Panel Data Developed and Applied to Test the Hypothesis of True State Dependence Against the Hypothesis of Spurious State Dependence," Annales de L'INSEE Paris, 30-31, 227-270.

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

Heckman, J., (1981a), "Heterogeneity and State Dependence," in Studies in Labor Markets, S. Rosen ed., Chicago: University of Chicago Press.

Heckman, J., (1981b), "Statistical Models for Discrete Panel Data," in Structural Analysis of Discrete Data with Econometric Applications, C. Manski and D. McFadden eds., Cambridge: MIT Press.

Heckman, J., (1996), "Randomization as an Instrumental Variable," Review of Economics and Statistics, 73, 336-340.

Heckman, J., (1997), "Instrumental Variables," Journal of Human Resources, 32, 441-462.

Heckman, J., H. Ichimura, J. Smith, and P. Todd, (1998), "Characterizing Selection Bias Using Experimental Data," Econometrica, 66, 1017-1098.

Heckman, J., H. Ichimura, and P. Todd, (1998), "Matching as an Econometric Evaluation Estimator," Review of Economic Studies, 65, 261-294.

Heckman, J. and D. Neal, (1996), "Coleman's Contributions to Education: Theory, Research Styles and Empirical Research," in James S. Coleman, ed. J. Clark, Washington D.C: Falmer Press.

Heckman, J. and R. Robb, (1985), "Alternative Methods for Evaluating the Impact of Interventions," in Longitudinal Analysis of Labor Market Data, J. Heckman and B. Singer, eds., New York: Cambridge University Press.

Heckman, J. and B. Singer, (1984a), "Econometric Duration Analysis," Journal of Econometrics, 24, 63-132.

Heckman, J. and B. Singer, (1984b), "A Method for Minimizing the Impact of

Distributional Assumptions in Econometric Models for Duration Data," Econometrica, 52, 271-320.

Heckman, J. and B. Singer, (1984c), "The Identifiability of the Proportional Hazards Model," Review of Economic Studies, 51, 231-241.

Heckman, J. and B. Singer, (1985), "Social Science Duration Analysis," in Longitudinal Analysis of Labor Market Data, J. Heckman and B. Singer, eds., New York: Cambridge University Press.

Heckman, J., and J. Smith, (1995), "Assessing the Case for Social Experiments," Journal of Economic Perspectives, 9, 83-110.

Henderson, J. V., R. Mieszkowski, and Y. Sauvageau, (1978), "Peer Group Effects and Educational Production Functions," Journal of Public Economics, 10, 97-106.

Hoffman, S. and R. Plotnick, (1996), "The Effect of Neighborhood Characteristics on Young Adult Outcomes: Alternative Estimates," Institute for Research on Poverty, University of Wisconsin, Discussion Paper no. 1106-96.

Honoré, B. and E. Kyriazidou, (1998), "Panel Data Discrete Choice Models with Lagged Dependent Variables," mimeo, Princeton University.

Hsiao, C., (1983), "Identification," in Handbook of Econometrics 1, Z. Griliches and M. Intriligator, eds., Amsterdam: North Holland.

Ioannides, Y., (1990), "Trading Uncertainty and Market Structure," International Economic Review, 31, 619-638.

Ioannides, Y., (1997a), "Evolution of Trading Structures," in The Economy as a Complex Evolving System II, W. B. Arthur, S. Durlauf and D. Lane, eds. Redwood City: Addison-Wesley.

Ioannides, Y. (1997b), "Nonlinear Neighborhood Interactions and Intergenerational Human Capital," working paper, Tufts University.

Ioannides, Y., (1999), "Residential Neighborhood Effects," working paper, Tufts University.

Irwin, E. and N. Bockstael, (1998), "Interacting Agents, Spatial Externalities, and the Endogenous Evolution of Land Pattern Use," working paper no. 98-24, Department of Agricultural and Resource Economics, University of Maryland.

Jencks, C. and S. Mayer, (1990), "The Social Consequences of Growing Up in a Poor Neighborhood," in Inner-City Poverty in the United States, L. Lynn and M.

McGreary, eds., Washington D.C.: National Academy Press.

Jones, A., (1994), "Health, Addiction, Social Interaction, and the Decision to Quit Smoking," Journal of Health Economics, 13, 93-110.

Jones, S., (1984), The Economics of Conformism, Oxford: Basil Blackwell.

Jovanovic, B., (1987), "Micro Shocks and Aggregate Risk," Quarterly Journal of Economics, CII, 395-409.

Jovanovic, B., (1989), "Observable Implications of Models with Multiple Equilibria," Econometrica, 57, 1431-1439.

Kac, M., (1968), "Mathematical Mechanisms of Phase Transitions," in Statistical Physics: Phase Transitions and Superfluidity, vol. 1, M. Chretien, E. Gross, and S. Desar eds., Brandeis University Summer Institute in Theoretical Physics 1966.

Kapetyn, A., S. van de Geer, H. van de Stadt, and T. Wansbeek, (1997), "Interdependent Preferences: An Econometric Analysis," Journal of Applied Econometrics, 12, 665-686.

Katz, L., J. Kling and J. Liebman, (1997), Moving to Opportunity in Boston: Early Impacts of a Housing Mobility Program, mimeo, Harvard University.

Katz, M. and C. Shapiro, (1986), "Technology Adoption in the Presence of Network Externalities," Journal of Political Economy, 94, 822-841.

Kaufman, J. and J. Rosenbaum, "The Education and Employment of Low-Income Black Youth in White Suburbs," Educational Evaluation and Policy Analysis, 14, 3, 229-240.

Kelly, M., (1997), "The Dynamics of Smithian Growth," Quarterly Journal of Economics, CXII, 939-964.

King, G., (1997), A Solution to the Ecological Inference Problem, Princeton: Princeton University Press.

Kirman, A., (1983), "Communication in Markets: A Suggested Approach," Economic Letters, 12, 1-5.

Kirman, A., (1997), "The Economy as an Interactive System," in The Economy as a Complex Evolving System II, W. B. Arthur, S. Durlauf and D. Lane, eds. Redwood City: Addison-Wesley.

Kirman, A., C. Oddou, and S. Weber, (1986), "Stochastic Communication and

Coalition Formation," Econometrica, 54, 129-138.

Kitcher, P., (1993), The Advancement of Science, Oxford: Oxford University Press.

Kohler, H.-P., (1997), Fertility and Social Interaction: An Economic Approach, PhD Dissertation, Department of Economics, University of California, Berkeley.

Kollman, K., J. Miller, and S. Page, (1992), "Adaptive Parties in Spatial Elections," American Political Science Review, 86, 929-937.

Kollman, K., J. Miller, and S. Page, (1997a), "Political Parties and Electoral Landscape," British Journal of Political Science,

Kollman, K., J. Miller, and S. Page, (1997b), "Computational Political Economy," in The Economy as a Complex Evolving System II, W. B. Arthur, S. Durlauf and D. Lane, eds. Redwood City: Addison-Wesley.

Krauth, B., (1998), "Social Interactions, Thresholds, and Unemployment in Neighborhoods," mimeo, University of Wisconsin.

Kremer, M., (1997), "How Much Does Sorting Increase Inequality?," Quarterly Journal of Economics, CXII, 115-140.

Krosnick, J. and C. Judd, (1982), "Transitions in Social Influence in Adolescence: Who Induces Cigarette Smoking," Developmental Psychology, 81, 359-368.

Krugman, P., (1996), The Self-Organizing Economy, Oxford: Basil Blackwell.

Kuhn, T., (1970), The Structure of Scientific Revolutions, second edition, Chicago: University of Chicago Press.

Kyriazidou, E., (1997a), "Estimation of a Panel Data Sample Selection Model," Econometrica, 65, 1335-1364.

Kyriazidou, E., (1997b), "Estimation of Dynamic Panel Data Sample Selection Model," mimeo, University of Chicago.

Labov, W., (1972a), Sociolinguistic Patterns, Philadelphia, University of Pennsylvania Press.

Labov, W., (1972b,), Language in the Inner City, Philadelphia, University of Pennsylvania Press.

Ladd, H. and J. Ludwig, (1998), "The Effects of MTO on Educational

Opportunities in Baltimore: Early Evidence," working paper, Joint Center for Poverty Research, Northwestern University.

Lancaster, T., (1990), The Analysis of Transition Data, New York: Cambridge University Press.

Leamer, E., (1983), "Let's Take the Con out of Econometrics," American Economic Review, 73, 31-43.

Legros, P. and A. Newman, (1997), "Matching in Perfect and Imperfect Worlds," working paper, Columbia University.

Liggett, T., (1985), Interacting Particle Systems, New York: Springer-Verlag.

Lindbeck, A., S. Nyberg, and J. Weibull, (1999), "Social Norms and Economic Incentives in the Welfare State," Quarterly Journal of Economics, CXIV, 1-36.

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

Lukacs, E., (1975), Stochastic Convergence, New York: Academic Press.

Manski, C., (1975), "The Maximum Score Estimation of the Stochastic Utility Model of Choice," Journal of Econometrics, 3, 205-228.

Manski, C. and D. McFadden, (1981), "Econometric Models of Probabilistic Choice," in Structural Models of Discrete Data with Econometric Applications, Cambridge: MIT Press.

Manski, C., (1985), "Semiparametric Estimation of Discrete Response: Asymptotic Properties of the Maximum Score Estimator," Journal of Econometrics, 27, 313-333.

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

Manski, C., (1993a), "Identification Problems in the Social Sciences," in Sociological Methodology, vol. 23, P. Marsden, ed., Cambridge: Basil Blackwell.

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

Manski, C., (1995), Identification Problems in the Social Sciences, Cambridge: Harvard University Press.

Manski, C., (1997), "Identification of Anonymous Endogenous Interactions," in The Economy as a Complex Evolving System II, W. B. Arthur, S. Durlauf and D. Lane, eds. Redwood City: Addison-Wesley.

Manski, C., and J. Pepper, (1998), "Monotone Instrumental Variables: With an Application to the Returns on Schooling, mimeo, Northwestern University.

Mayer, S., (1991), "How Much Does a High School's Racial and Socioeconomic Mix Affect Graduation and Teenage Fertility Rates?," in The Urban Underclass, C. Jencks and P. Peterson eds., Washington DC: Brookings Institution Press.

McFadden, D., (1974), "Conditional Logit Analysis of Qualitative Choice Behavior," in P. Zarembka, Frontiers in Econometrics, New York: Academic Press.

McFadden, D., (1984), "Econometric Analysis of Qualitative Response Models," Handbook of Econometrics, Volume II, Z. Griliches and M. Intrilligator eds., Amsterdam: North-Holland.

McManus, D., (1992), "How Common is Identification in Parametric Models?," Journal of Econometrics, 53, 5-23.

Milgrom, P., and D. J. Roberts, (1990), "Rationalizability, Learning, and Equilibrium in Games with Strategic Complementarities," Econometrica, 58, 1255-1277.

Miyao, T., (1978b), "A Probabilistic Model of Location Choice with Neighborhood Effects," Journal of Economic Theory, 19, 34-58.

Moffitt, R., (1998), "Policy Interventions, Low-Level Equilibria, and Social Interactions," mimeo, Johns Hopkins University.

Montgomery, J., (1991), "Social Networks and Labor-Market Analysis: Towards an Economic Analysis," American Economic Review, 81, 1408-1418.

Montgomery, J., (1992), "Social Networks and Persistent Inequality in Labor Markets," mimeo, Northwestern University.

Morris, S., (1998), "Contagion," mimeo, University of Pennsylvania.

Munshi, K. and J. Myaux, (1998), "Social Effects in the Demographic Transition: Evidence from Matlab, Bangladesh" mimeo, Boston University.

Murray, J., (1984), Asymptotic Analysis, New York: Springer-Verlag.

Neal, D., (1997), "The Effects of Catholic Secondary Schooling on Educational Achievement," Journal of Labor Economics, 15, 98-123.

Nechyba, T., (1996), "Social Approval, Values and AFDC: A Re-examination of the Illegitimacy Debate," mimeo, Stanford University.

Oomes, N., (1998), "Market Failures in the Economics of Science: Barriers to Entry, Increasing Returns, and Lock-In by Historical Events," mimeo, University of Wisconsin at Madison.

Page, S., (1997), "Network Structure Matters," working paper, University of Iowa.

Pakes, A., (1999), "A Framework for Applied Dynamic Analysis in I. O.," mimeo, Department of Economics, Harvard University.

Popkin, S., J. Rosenbaum, and P. Meaden, (1993), "Labor Market Experiences of Low-Income Black Women in Middle-Class Suburbs: Evidence from a Survey of Gautreaux Program Participants," Journal of Policy Analysis and Management, 12, 556-574.

Postlewaite, A., (1997), "The Social Basis of Interdependent Preferences," mimeo, University of Pennsylvania.

Powell, J., (1987), "Semiparametric Estimation of Bivariate Latent Variable Models," working paper, Social Systems Research Institute, University of Wisconsin.

Raftery, A., (1995), "Bayesian Model Selection in Social Science Research," Sociological Methodology, P. Marsden, ed., Oxford: Blackwell.

Raftery, A., D. Madigan, and J. Hoeting, (1997), "Bayesian Model Averaging for Linear Regression Models," Journal of the American Statistical Association, 92, 179-191.

Rivkin, S., (1997), "The Estimation of Peer Group Effects," mimeo, Amherst College.

Romer, D., (1993), "Rational Asset-Price Movements Without News," American Economic Review, 83, 1112-1130.

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.

Rosenbaum, J., S. DeLuca, and S. Miller, (1999), "The Long-Term Effects of Residential Mobility on AFDC Receipt: Studying the Gautreaux Program with Administrative Data," mimeo, Northwestern University.

Rosser, J. B., (1999), "On the Complexities of Complex Economic Dynamics," Journal of Economic Perspectives, forthcoming.

Rothenberg, T., (1971), "Identification in Parametric Models," Econometrica, 39, 577-591.

Rouse, C., (1998a), "Private School Vouchers and Student Achievement: An Evaluation of the Milwaukee Parental Choice Program," Quarterly Journal of Economics, 553-602.

Rouse, C., (1998b), "Schools and Student Achievement: More Evidence from the Milwaukee Parental Choice Program," working paper #396, Industrial Relations Section, Princeton University.

Sah, R., (1991), "Social Osmosis and Crime," Journal of Political Economy, 99, 1272-1295.

Sampson, R. and J. Laub, (1995), Crime in the Making, Harvard: Harvard University Press.

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

Sattinger, M., (1975), "Comparative Advantage and the Distribution of Earnings," Econometrica, 43, 455-468.

Schelling, T., (1971), "Dynamic Models of Segregation," Journal of Mathematical Sociology, 1, 143-186.

Schelling, T., (1973), "Hockey Helmets, Concealed Weapons, and Daylight Saving: A Study of Binary Choices with Externalities," Journal of Conflict Resolution, 17, 381-428.

Shnirman, N., (1968) "On the Ergodicity of a Markov Chain," Problems in Information Theory, 20, 115-124.

Sims, C., (1980), "Macroeconomics and Reality," Econometrica, 48, 1-48.

Slavin, R., (1990), "Achievement Effects of Ability Grouping in Secondary Schools: A Best-Evidence Synthesis," Review of Educational Research, 60, 471-499.

Solon, G., M. Page, and G. Duncan, (1999), "Correlations Between Neighboring Children in Their Subsequent Educational Attainment," mimeo, University of Michigan.

Spitzer, F., (1971a), "Markov Random Fields and Gibbs Ensembles," American Mathematical Monthly, 78, 142-154.

Spitzer, F., (1971b), Random Fields and Interacting Particle Systems, Providence: Mathematical Association of America.

Stavskaya, O. and I. Pyatetskii-Shapiro, (1968), "Homogeneous Networks of Spontaneous Active Elements," Problems in Information Theory, 20, 91-106.

Steinberg, L. with B. Brown and S. Dornbusch, (1996), Beyond the Classroom, New York: Simon and Schuster.

Streufert, P., (1991), The Effect of Underclass Isolation on Schooling Choice," mimeo, University of Wisconsin at Madison.

Sucoff, C. and D. Upchurch, (1998), "Neighborhood Context and the Risk of Childbearing Among Metropolitan-Area Black Adolescents," American Sociological Review, 63, 571-585.

Topa, G., (1997), "Social Interactions, Local Spillovers and Unemployment," mimeo, New York University.

Verbrugge, R., (1999), "A Framework for Studying Economic Interactions (with Applications to Corruption and Business Cycles," mimeo, Department of Economics, Virginia Polytechnic Institute.

Wallis, K., (1980), "Econometric Implications of the Rational Expectations Hypothesis," Econometrica, 48, 49-73.

Weisberg, B., P. Reagan, and J. Yankow, (1999), "Do Neighborhoods Affect Work Behavior? Evidence from the NLSY79," mimeo, Department of Economics, Ohio State University.

Wilson, W. J., (1987), The Truly Disadvantaged, Chicago: University of Chicago Press.

Witte, J., (1997), "Achievement Effects of the Milwaukee Voucher Program," mimeo, Department of Political Science, University of Wisconsin.

Yinger, J., (1995), Closed Doors, Opportunities Lost, New York: Russell Sage Foundation.

Young, H. P., (1998), Individual Strategy and Social Structure, Princeton: Princeton University Press.

Zax, J., and D. Rees, (1998), "Environment, Ability, Effort, and Earnings," working paper, no. 9801, Center for Research on Economic and Social Policy, University of Colorado at Denver.