9217

DI

University of Wisconsin - Madison, Social Systems research institutes

SSRI workshop series

Identification Problems in the Social Sciences

9217

Charles F. Manski

IDENTIFICATION PROBLEMS IN THE SOCIAL SCIENCES

Charles F. Manski
Department of Economics
University of Wisconsin
Madison, WI 53706

August 1992

ABSTRACT

Methodological research in the social sciences aims to learn what conclusions can and cannot be drawn given empirically relevant combinations of assumptions and data. Methodologists have long found it useful to separate inferential problems into statistical and identification components. Studies of identification seek to characterize the conclusions that could be drawn if the researcher had available a sample of unlimited size. Studies of statistical inference seek to characterize the generally weaker conclusions that can be drawn given a sample of positive but finite size. Statistical and identification problems limit in distinct ways the conclusions that may be drawn in empirical research. Statistical problems are most severe when the available sample is small. Identification problems are most severe when the researcher knows little about the population under study and the sampling process yields only weak data on the population. This paper synthesizes some of my recent research and thinking on identification problems in the social sciences. Four problems are discussed: extrapolation of regressions, the selection problem, identification of endogenous social effects from outcome data, and identification of subjective phenomena. These problems arise regularly in social science research and are the source of many substantive disputes.

Invited Paper, Section on Methodology, 1992 Annual Meeting of the American Sociological Association.

CONTENTS

1. Introduction

The members of our open society often express differing views on social policy. Disagreements presumably arise out of the conflicting self-interests and ideologies of the population, but normative differences are not the only contributing force. Many controversies reflect divergent beliefs about human behavior, specifically about the effects of government programs on behavior.

Consider, for example, the continuing debate about welfare. Perspectives on AFDC and other welfare programs appear in part to reflect beliefs about the way these programs affect marriage, fertility, and labor supply behavior. Almost everyone has an opinion on the matter but the opinions vary widely.

Divergent beliefs about human behavior should, one might think, be reconciliable through empirical social science research. Yet social scientists rarely seem able to settle questions of public concern. During the past twenty years researchers have worked hard to learn how welfare affects behavior (see Moffitt, 1992) and to evaluate the job training programs that aim to move welfare recipients into the labor market (see Manski and Garfinkel, 1992). We have similarly worked to understand how neighborhoods influence their inhabitants (see Jencks and Mayer, 1989), how the threat of punishment deters crime (see Blumstein et al., 1978), how school attributes affect student learning (see Hanushek, 1986, and Gamoran, 1992), and how early childbirth affects the lives of mothers and their children (see Hayes and Hofferth, 1987). In these and so many other areas, progress seems painfully slow.

Indeed, the cumulative research on a subject only rarely converges toward a consensus.

Why has empirical research in the social sciences so often failed to yield clearcut answers to questions of interest? The social sciences may be immature. Or it may just be hard to answer the questions that the social sciences are asked to address.

I believe the core problem to be the inherent difficulty of the questions facing the social sciences. The conclusions that one can draw from an empirical analysis are determined by the assumptions and the data that one brings to bear. In social science research, the available data are typically limited and the range of plausible assumptions wide; hence the generally accepted conclusions are necessarily weak. Disagreements about the determinants of human behavior, the nature of social interactions, and the consequences of public policy persist because researchers who analyze the same data under different maintained assumptions reach different logically valid conclusions.

Although the core problem of the social sciences is the difficulty of the enterprise, there is also a problem of immaturity. Many social scientists do not appreciate the core problem. Some seem not to recognize that the interpretation of data requires assumptions. How often do we see an empirical analysis that applies some conventional statistical method with little understanding of the assumptions needed to interpret the results in the conventional way? Some researchers understand the logic of scientific inference but nevertheless deny it when reporting their own work. The

scientific community rewards researchers who produce strong novel findings and the public, impatient for solutions to its pressing policy concerns, rewards researchers who offer simple analyses leading to unequivocal policy recommendations. These incentives make it tempting for researchers to maintain assumptions far stronger than they can persuasively defend, in order to draw strong conclusions. When this happens, empirical research degenerates into the advocacy of "forensic" social science, where researchers sharing the same data but maintaining different assumptions argue about the interpretation of the data. With empirical resolution impossible, scientific inquiry is replaced by debate.

STATISTICAL INFERENCE AND IDENTIFICATION: Methodological research in the social sciences aims to learn what conclusions can and cannot be drawn given empirically relevant combinations of assumptions and data. For at least a century, methodologists have used statistical theory to frame their studies (see Stigler, 1986, and Clogg, 1992). One supposes that the empirical problem is to infer some feature of a population described by a probability distribution and that the available data are observations extracted from the population by some sampling process. One combines the data with assumptions about the population and the sampling process to draw statistical conclusions about the population feature of interest.

Working within this familiar framework, methodologists have found it useful to separate the inferential problem into statis-

tical and identification components. Studies of identification seek to characterize the conclusions that could be drawn if the researcher had available a sample of unlimited size. Studies of statistical inference seek to characterize the generally weaker conclusions that can be drawn given a sample of positive but finite size. Statistical and identification problems limit in distinct ways the conclusions that may be drawn in empirical research. Statistical problems are most severe when the available sample is small. Identification problems are most severe when the researcher knows little about the population under study and the sampling process yields only weak data on the population.

Statistical problems contribute to the difficulty of empirical research but identification is the core problem of the social sciences. Increasing the sizes of our available data samples would enable us to sharpen the inferences we now make but would not enable us to make new kinds of inferences. New inferences require either new knowledge of the population under study or new sampling processes yielding data on different features of the population.

FOCUS ON IDENTIFICATION: Beginning in the early 1980s I have, in my research and teaching, gradually devoted less time to the study of statistical questions and more to the analysis of identification. I now find it natural to study inference in two stages. First one determines what restrictions on the population of interest are implied by the available prior information and by the sampling process generating data. Then one develops methods for estimating

identified population features, usually by treating the sample as analogous to the population (see Manski, 1988a). Both stages in the study of inference are important, but the first is more fundamental and more in need of fresh thinking.

This paper synthesizes some of what I have learned about identification problems in the social sciences. Early on, I found that effective study of identification requires an appropriate balance between generality and specificity. An overly general analysis may yield only sterile theorems stating that a given population feature is identified if some system of equations or extremum problem has a unique solution. An overly specific analysis may obscure basic ideas. With these concerns in mind, I have chosen to discuss four identification problems: extrapolation of regressions (Section 2); the selection problem (Section 3); identification of endogenous social effects from outcome data (Section 4); and identification of subjective phenomena (Section 5). These four problems arise regularly in social science research and are the source of many substantive disputes.

I would like to call the reader's attention to several themes that arise in the course of considering these identification problems:

* A fruitful approach to the study of identification is to begin by asking what can be learned from the data alone, in the absence of prior information. Once this is established, one then asks what more may be learned given various types of prior information (see Sections 2 and 3).

* It may be easier to identify some population features than others. For example, the median is easier to identify than the mean when the data are censored (see Section 3).

* Identification is not an all-or-nothing proposition. One may not have enough information to learn the value of a parameter, but may be able to bound it (see Sections 3 and 5).

* Outcome data alone reveal little about the channels through which society influences the individual (see Section 4).

2. Extrapolation of Regressions

The problem of extrapolating regressions is very familiar and so forms a good starting point for our discussions. Consideration of extrapolation also serves to introduce basic ideas of nonparametric regression analysis used repeatedly in Sections 3 through 5.

Informally, extrapolation is prediction of a variable y given a specified value for another variable x, in the absence of data on the behavior of y when x takes this value. Formally, let $Y \times X$ be the space of logically possible values of (y,x). Assume there is a probability distribution on $Y \times X$ and that a random sampling process yields observations of (y,x). Suppose that, given a value x_0 in X, one wishes to make the best prediction of y, in the sense of minimizing square loss. As is well-known, the best predictor in this sense is $E(y|x)$, the mean regression of y on x. Extrapolation is the problem of identifying $E(y|x=x_0)$ when the regressor x_0 is logically possible but is off the support of x. (The point x_0^* is on the support of x if there is positive probability of observing x arbitrarily close to x_0 and is off the support if there is zero probability of observing x within some neighborhood of x_0.)

Identification of $E(y|x)$ on and off the support of x present very different challenges. Minimal prior information about the population suffices to identify the regression on the support; indeed, the literature on nonparametric regression analysis shows that it is easy to estimate $E(y|x)$ on the support.[1] On the other

hand, extrapolation requires substantial prior information. These matters are explained in Sections 2.1 and 2.2.

2.1. IDENTIFICATION AND CONSISTENT ESTIMATION ON THE SUPPORT

There are two cases to consider. Suppose first that the point x_0 is not only on the support but that $\text{Prob}(x=x_0) > 0$. Then $E(y|x=x_0)$ is identified and can be estimated consistently given only the assumption that $E(y|x=x_0)$ exists and is finite. An obvious estimate is the sample average of y across the observations for which $x_i = x_0$, namely

$$(2.1) \qquad \frac{\sum_{i=1}^{N} y_i \, 1[x_i=x_0]}{\sum_{j=1}^{N} 1[x_j=x_0]} .$$

(The indicator function 1[.] takes the value one if the bracketed logical condition holds and zero otherwise.) The strong law of large numbers implies that the cell average (2.1) is a consistent estimate of the conditional mean $E(y|x=x_0)$.

Now suppose that x_0 is on the support but that $\text{Prob}(x=x_0) = 0$. This is the situation when x has a continuous distribution with positive density at x_0. The cell-average estimate no longer works; with probability one, the event $x_i = x_0$ never occurs in the sample. On the other hand, one can estimate $E(y|x=x_0)$ by the sample average

of y across those observations for which x_i is suitably near x_0; that is, by a "local average" of the form

$$(2.2) \qquad \frac{\sum_{i=1}^{N} y_i \, \rho(x_i, x_0) < \delta_N]}{\sum_{j=1}^{N} 1[\rho(x_j, x_0) < \delta_N]} ,$$

Here $\rho(.,.)$ is any sensible metric measuring the distance between x_0 and x_i; for example, Euclidean distance will do. The parameter δ_N is a sample-size dependent "bandwidth" chosen by the researcher to operationalize the idea that one wishes to average over those observations in which x_i is near x_0.

This simple nonparametric approach to estimation of $E(y|x=x_0)$ works, in the sense of providing a consistent estimate, if

(a) $E(y|x)$ is continuous at $x = x_0$ and $Var(y)$ exists.

(b) one tightens the bandwidth as the sample size increases.

(c) one does not tighten the bandwidth too rapidly.

Of these conditions, only (a) requires prior information about the population and the required information is very weak indeed.

To understand why conditions (a), (b), and (c) suffice, suppose that δ_N is kept fixed at some value δ. Then as N increases, the strong law of large numbers implies that the estimate (2.2) converges to $E[y|\rho(x_i, x_0) < \delta]$; that is, to the mean of y conditional on x being within δ of x_0. If $E(y|x)$ is continuous at $x = x_0$, then as δ approaches zero, $E[y|\rho(x, x_0) < \delta]$ approaches $E(y|x=x_0)$. These two facts suggest that an estimate converging to $E(y|x=x_0)$ can be

obtained by adopting a bandwidth-selection rule that makes δ_N approach zero as N increases. It can be shown that this heuristic idea succeeds provided that the variance of y exists and that δ_N does not approach zero too quickly. In particular, the rate at which δ_N approaches zero must be slower than $1/N^{1/K}$, where K is the dimension of the vector x. This condition ensures that the number of observations actually used to calculate the estimate (2.2) increases with the sample size N.[2]

2.2. IDENTIFICATION OFF THE SUPPORT

The local-average estimate (2.2) does not work when x_0 is off the support of x. Suppose that there is zero probability of observing x within some distance d_0 of x_0. Then the estimate (2.2) ceases to exist when one attempts to reduce the bandwidth δ_N below d_0.

The failure of the local-average estimate is symptomatic of a fundamental problem: in the absence of prior information, the distribution of y conditional on x_0 is not identified when x_0 is off the support of x. The random sampling process alone identifies the joint distribution of (y,x) but no more. When x_0 is off the support, changing the distribution of y conditional on x_0 has no effect on the joint distribution of (y,x).

2.2.1. Local Smoothness Assumptions

What kind of prior information does and does not identify the regression off the support? An important negative fact is that local smoothness assumptions on $E(y|x)$ do not suffice. Suppose that $Var(y)$ exists and that $E(y|x)$ is continuous on all of X. Then $E(y|x=x_1)$ is identified and can be consistently estimated at every point x_1 on the support of x. But we have no information about the value of $E(y|x=x_0)$ at points x_0 off the support.

To understand why, let x_1 be the point on the support that is closest to x_0. A continuity assumption implies that $E(y|x)$ is near $E(y|x=x_1)$ when x is near x_1, but does not tell us how to interpret the two uses of the word "near" as magnitudes. In particular, we do not know whether the distance separating x_0 and x_1 should be interpreted as large or small.

2.2.2. Global Restrictions

Identification off the support requires prior information that restricts the regression globally rather than locally. The traditional practice has been to assert a parametric model for $E(y|x)$, for example a linear model

$$(2.3) \quad E(y|x) = x'b.$$

Suppose that the components of x are linearly independent, so that the parameter vector b is identified. Then (2.3) may be applied to identify $E(y|x)$ at all logically possible values of x, whether on

or off the support. Weaker global restrictions allowing partial extrapolation appear in the literature on semiparametric regression analysis (see Manski,1988b).

The problem with global restrictions, of course, is that the assumptions made about the form of the regression may be wrong. A model may fail either on or off the support of x. Failure on the support is detectable. The classical statistical theory of hypothesis testing was developed for just this purpose. Failure off the support is a qualitatively different problem as it is inherently not detectable.

Irresolvable disagreements arise when researchers hypothesize models that agree with $E(y|x)$ on the support of x but that behave differently off the support. Given a specified sampling process, there is no empirical way to discriminate among models all of which "fit the data." The only ways to judge the extrapolations implied by such models are by subjectively assessing the plausibility of the models or by initiating a new sampling process that gathers data at values of x where the various models yield different values for $E(y|x)$.

2.3. IDENTIFICATION OF CONTRASTS

One often wants to contrast the regression at two values of x. Then the object of interest is $E(y|x=x_1) - E(y|x=x_0)$ for specified x_0 and x_1. This contrast can sometimes be interpreted as the

"effect" on y of changing the regressor from x_0 to x_1 (see Section 3.4.2).

The discussion of Section 2.1 applies if both x_0 and x_1 are on the support of x. Otherwise, the discussion of Section 2.2 applies to either or both of x_0 or x_1, as the case may be. There is little else of a general nature to say about identification of contrasts, but I would like to call attention to a familiar problem that arises when the vector of regressors is functionally dependent. An instance of this problem will be seen in Section 4.

Let $x = (w,v)$, where w are v are vectors. One is often interested in a contrast of the form $E(y|w=w_1,v) - E(y|w=w_0,v)$. That is, v is held fixed and w is varied between the values w_0 and w_1. Suppose that the regressor values (w_0,v) and (w_1,v) are both logically possible but that w happens to be a function of v within the population; say $w = f(v)$. Then (w_0,v) and (w_1,v) cannot both be on the support of x; at most, $[f(v),v]$ is on the support.

Thus, functional dependence implies that identification of the contrast $E(y|w=w_1,v) - E(y|w=w_0,v)$ requires global restrictions on the regression. One might, for example, know that $E(y|w,v)$ is linear in (w,v). This information identifies $E(y|w=w_1,v) - E(y|w=w_0,v)$ provided that the function f(.) is not linear.

3. The Selection Problem

Some respondents to a household survey decline to report their incomes. Some women responding to a longitudinal fertility survey complete their childbearing after the survey is terminated. Some welfare recipients do not enroll in a job training program. These very different situations share the common feature that a variable is censored; respondents' incomes, women's completed family sizes, or welfare recipients' employment status after job training.

Social scientists constantly seek to draw conclusions from censored data. We routinely pose and try to answer questions of the form:

What is the effect of ____ on ____?

For example,

What is the effect of the AFDC program on labor supply?

What is the effect of schooling on wages?

What is the effect of family structure on children's outcomes?

All efforts to address such "treatment effect" questions must confront the fact that the data are inherently censored. One wants to compare outcomes across different treatments but each unit of analysis, whether a survey respondent or experimental subject, experiences only one of the treatments under consideration.

Whereas the implications of censoring were not well appreciated twenty years ago, they are much better understood today. In particular, methodologists have devoted substantial attention to the selection problem: the problem of identifying regressions from

random samples in which the realizations of regressors are always observed but the realizations of outcomes are censored. The selection problem is logically separate from the extrapolation problem discussed in Section 2. The extrapolation problem follows from the fact that a random sampling process does not yield observations of y off the support of x. The selection problem arises when a censored random sampling process does not fully reveal the behavior of y on the support of x. So selection presents new challenges on top of those faced in extrapolation.

NATURE OF THE PROBLEM: To introduce the selection problem formally, suppose that each member of the population is characterized by a triple (y,z,x). Here y is scalar, x is a vector, and z is a binary variable taking the value 0 or 1.[3] One draws a random sample from the population and observes all the realizations of (z,x), but observes y only when $z = 1$. One wants to learn some feature of the probability distribution of y conditional on x, denoted $P(y|x)$; that is, one wants to learn some regression of y on x.

The problem is the failure of the censored-sampling process to identify $P(y|x)$ on the support of x. To isolate the difficulty, decompose $P(y|x)$ into the sum

$$(3.1) \quad P(y|x) = P(y|x,z=1)P(z=1|x) + P(y|x,z=0)P(z=0|x).$$

The sampling process identifies the selection probability $P(z=1|x)$, the censoring probability $P(z=0|x)$, and the distribution of y

conditional on selection, $P(y|x,z=1)$. It is uninformative regarding the distribution of y conditional on censoring, $P(y|x,z=0)$. Hence the censored-sampling process reveals only that

$$(3.2) \quad P(y|x) \in [P(y|x,z=1)P(z=1|x) + \gamma P(z=0|x),\ \gamma \in \Gamma],$$

where Γ denotes the space of all probability distributions on the real line.

The logical starting point for investigation of the selection problem is to characterize the problem in the absence of prior information; that is, to learn what restrictions on $P(y|x)$ are implied by (3.2) alone. Section 3.1 summarizes my recent work on this subject. Then Sections 3.2 and 3.3 explore identification when prior information and/or richer data are available. Section 3.2 considers various types of prior information brought to bear in the econometric and statistical literatures. Section 3.3 explains the additional identification possibilities that arise in the "switching regression" setting, where censoring of y is accompanied by observation of a different outcome s. Section 3.4 applies these findings to the important problem of identifying treatment effects.

3.1. IDENTIFICATION IN THE ABSENCE OF PRIOR INFORMATION

Inspection of (3.2) reveals that, in the absence of prior information on the distribution of (y,z,x), one cannot reject the conditional independence hypothesis

$$(3.3)\quad P(y|x) = P(y|x,z=1) = P(y|x,z=0).$$

Simply observe that (3.3) holds if one sets $\gamma = P(y|x,z=1)$ in (3.2). Econometricians usually refer to the conditional independence hypothesis as "exogenous" selection while some statisticians refer to it as "ignorable" selection.

In the absence of prior information, censoring makes it impossible to learn anything about the mean regression of y on x. To see this, decompose $E(y|x)$ into the sum

$$(3.4)\quad E(y|x) = E(y|x,z=1)P(z=1|x) + E(y|x,z=0)P(z=0|x).$$

The censored-sampling process identifies $E(y|x,z=1)$ and $P(z|x)$ but provides no information on $E(y|x,z=0)$, which might take any value below minus and plus infinity. Hence, whenever the censoring probability $P(z=0|x)$ is positive, the sampling process imposes no restrictions on $E(y|x)$.

These negative results do not, however, imply that the selection problem is fatal in the absence of prior information. In fact, censored data imply informative, easily interpretable bounds on many important features of the conditional distribution $P(y|x)$, including quantiles, probabilities, and the means of bounded functions of y. In what follows, I present some of the findings of Manski (1989,1992a).

3.1.1. Conditional Means of Bounded Functions of y

The central result, from which others may be derived, concerns the mean of a bounded function of y. Let g(.) be a real-valued function mapping y into a known bounded interval $[K_0,K_1]$, which may depend on g(.). Observe that

$$(3.5)\quad E[g(y)|x] = E[g(y)|x,z=1]P(z=1|x) + E[g(y)|x,z=0]P(z=0|x).$$

The sampling process identifies $E[g(y)|x,z=1]$ and $P(z|x)$ but provides no information on $E[g(y)|x,z=0)$. The last quantity, however, necessarily lies in the interval $[K_0,K_1]$. This simple fact yields the following:

$$(3.6)\quad E[g(y)|x,z=1]P(z=1|x) + K_0P(z=0|x) \le E[g(y)|x] \le E[g(y)|x,z=1]P(z=1|x) + K_1P(z=0|x).$$

Thus, a censored-sampling process bounds the mean regression of any bounded function of y. The lower bound is the value $E[g(y)|x]$ takes if, in the censored subpopulation, g(y) always equals K_0; the upper bound is the value of $E[g(y)|x]$ if all the censored y equal K_1. The bound is a proper subset of $[K_0,K_1]$, hence informative, whenever censoring is less than total. At each regressor value x_0, the bound width $(K_1-K_0)P(z=0|x=x_0)$ is proportional to the censoring probability $P(z=0|x=x_0)$. It is therefore meaningful to say that

the degree of underidentification of $E[g(y)|x=x_0]$ is proportional to the censoring probability at x_0.

3.1.2. Conditional Probabilities

The bound (3.6) has numerous applications. Perhaps the most farreaching is the bound it implies on the probability that y lies in any set $A \subset Y$. Let $g_A(.)$ be the indicator function $g_A(y) \equiv 1[y \in A]$. Then $E[g_A(y)|x] = P(y \in A|x)$, $K_0 = 0$, and $K_1 = 1$. Hence (3.5) implies that

$$(3.7)\quad P(y \in A|x,z=1)P(z=1|x) \leq P(y \in A|x) \leq P(y \in A|x,z=1)P(z=1|x) + P(z=0|x).$$

It is often convenient to characterize a probability distribution by its distribution function $P(y \leq t|x)$, $t \in R^1$. It follows from (3.7) that

$$(3.8)\quad P(y \leq t|x,z=1)P(z=1|x) \leq P(y \leq t|x) \leq P(y \leq t|x,z=1)P(z=1|x) + P(z=0|x).$$

It may seem surprising that one should be able to bound the distribution function of a random variable but not its mean. The explanation is a fact that is widely appreciated by researchers in the field of robust statistics: the mean of a random variable is not a continuous function of its distribution function. Hence

small perturbations in a distribution function can generate large movements in the mean. See Huber(1981).[4]

3.1.3. Conditional Quantiles

Let $\alpha \in (0,1)$. By definition, the α-quantile of y conditional on x is

$$(3.9) \quad q(\alpha,x) \equiv \min t: P(y \leq t | x) \geq \alpha.$$

In particular, the .5-quantile is the median. Interest in quantile regression analysis has developed rapidly over the past fifteen years, beginning from the work of Koenker and Bassett (1978). For an expository treatment, see Manski (1988a, Chapter 4).

The bound (3.8) on $P(y \leq . | x)$ can be inverted to show that $q(\alpha,x)$ must lie between two quantiles of the identified distribution $P(y|x,z=1)$. Define

$$(3.10) \quad r(\alpha,x) \equiv [1-(1-\alpha)/P(z=1|x)]\text{-quantile of } P(y|x,z=1)$$
$$\text{if } P(z=1|x) > 1-\alpha$$
$$\equiv -\infty \text{ otherwise.}$$
$$s(\alpha,x) \equiv [\alpha/P(z=1|x)]\text{-quantile of } P(y|x,z=1)$$
$$\text{if } P(z=1|x) \geq \alpha$$
$$\equiv \infty \text{ otherwise.}$$

It is proved in Manski (1992a) that

(3.11) $r(\alpha,x) \leq q(\alpha,x) \leq s(\alpha,x)$.

Moreover, in the absence of prior information, this bound on $q(\alpha,x)$ cannot be improved upon.

The lower and upper bounds $r(\alpha,x)$ and $s(\alpha,x)$ are increasing functions of α; hence the bound shifts to the right as α increases. The lower bound is informative if $P(z=1|x) > 1-\alpha$; the upper bound if $P(z=1|x) \geq \alpha$. So the bound (3.11) restricts $q(\alpha,x)$ to an interval of finite length if $P(z=1|x) > \max(\alpha,1-\alpha)$ and is uninformative if $P(z=1|x) < \min(\alpha,1-\alpha)$. In particular, the bound on the median regression is informative if $P(z=1|x) > 1/2$.

3.1.4. Sample Inference

The selection problem is, first and foremost, a failure of identification. It is only secondarily a difficulty in sample inference. To keep attention focussed on the central identification question, it is simplest to suppose that the conditional distributions identified by the sampling process, $P(y|x,z=1)$ and $P(z|x)$, are known. But it is also important to recognize that the population bounds reported above are easily estimable.

For example, estimation of the bound (3.6) is a conventional problem in nonparametric regression analysis of the type discussed in Section 2.1. Rewrite (3.6) in the equivalent form

(3.6') $E[g(y)z+K_{0g}(1-z)|x] \leq E[g(y)|x] \leq E[g(y)z+K_{1g}(1-z)|x]$.

The random variables $g(y)z+K_{0g}(1-z)$ and $g(y)z+K_{1g}(1-z)$ are both bounded; hence their variances exist. It follows that the lower and upper bounds $E[g(y)z+K_{0g}(1-z)|x]$ and $E[g(y)z+K_{1g}(1-z)|x]$ can be estimated consistently on the support of x as long as these quantities vary continuously in x. Given additional regularity conditions, asymptotically valid sampling confidence intervals can be placed around estimates of the bounds. An empirical application reporting bootstrapped confidence intervals is presented in Manski, Sandefur, McLanahan, and Powers (1992).

3.1.5. An Historical Note

It is of interest to ask why the simple bound results reported here took so long to appear. I believe that the explanation has at least three parts.

Timing has played a role. The modern literature on selection took shape in the 1970s, when the frontier of social science methodology was nonlinear parametric analysis. At that time, nonparametric regression analysis was just beginning to be developed by statisticians. Social scientists were not yet aware that nonparametric estimation of regressions was possible and did not think in the nonparametric terms needed to derive the bounds.

A second factor is the historical fixation of social scientists on point identification, which has inhibited appreciation of the potential usefulness of bounds. Estimable bounds on quantities that are not point-identified have been reported from time to time; a prominent early example appears in Frisch (1934). But the

conventional wisdom has been that bounds are hard to estimate and uninformative. Whatever the validity of this conventional wisdom in other contexts, it does not apply to the bound (3.6).

The preoccupation of researchers with the estimation of mean regressions has been a third factor. It has long been known that, in the absence of prior information, the selection problem is fatal for inference on the mean of an unbounded random variable. Social scientists have, improperly as it turns out, extrapolated that no inference at all is possible in the absence of prior information.

3.2. VARIETIES OF PRIOR INFORMATION

One can improve on the bounds reported in Section 3.1 if one possesses suitable prior information restricting the distribution of (y,z) conditional on x. A restriction has identifying power if it implies that $P(y|x)$ belongs to a set of distributions smaller than (3.2). Information restricting $P(y|x)$, $P(y|x,z=0)$, or $P(z|x,y)$ may have identifying power. Information restricting $P(y|x,z=1)$ or $P(z|x)$ is superfluous as the latter distributions are identified by the censored-sampling process alone.[5]

Ideally, we would like to learn the identifying power of all types of prior information, so as to characterize the entire spectrum of inferential possibilities. But there does not appear to be any effective way to conduct an exhaustive identification analysis. So researchers have investigated the power of specific bundles of restrictions thought to have application to empirical

problems of interest. Section 3.2.1 describes the latent-variable models developed by econometricians. Section 3.2.2 explains the quite different mixture-model approach favored by some statisticians. Section 3.2.3 presents my recent finding characterizing the identifying power of exclusion restrictions.

3.2.1. Econometric Latent Variable Models

Although the selection problem arises in many economic applications, econometricians have analyzed the problem in a sustained way only since the early 1970s. Before then, researchers generally maintained the exogenous-selection hypothesis (3.3), a notable exception being Tobin (1958).

The empirical plausibility of (3.3) was eventually questioned sharply. In particular, researchers observed that in many economic settings, the process by which observations on y become censored is related to the value of y (see Gronau, 1974). It also became clear that exogenous selection is not necessary to identify $P(y|x)$. An alternative is to specify a latent-variable model jointly explaining (y,z) conditional on x. (See, for example, Heckman, 1976; Maddala, 1983; or Winship and Mare, 1992).

For the past twenty years, econometric thinking on the selection problem has been expressed primarily through latent-variable models of the form

$$(3.12a) \quad y = f_1(x) + u_1$$

$$(3.12b) \quad z = 1[f_2(x) + u_2 > 0].$$

Here $[f_1(.), f_2(.)]$ are real functions of x and (u_1, u_2) are random variables whose realizations are unobserved by the researcher. The threshold-crossing form of the selection function (3.12b) is well-motivated in empirical analyses where the observability of y is determined by the binary choice behavior of a rational decision maker. In such cases $f_2(x)+u_2$ is the difference between the values of the two alternatives and (3.12b) states that the more highly-valued alternative is chosen.

Equations (3.12) alone do not restrict the distribution of (y,z) conditional on x. A model takes on content when restrictions are imposed on $[f_1(.), f_2(.)]$ and on the distribution of (u_1, u_2) conditional on x. The overriding concern of the literature has been to find plausible restrictions that identify the mean regression of y on x, although most of the restrictions studied actually identify the conditional distribution $P(y|x)$ fully. In what follows, I describe three types of restrictions that have received considerable attention. These restrictions are neither nested nor mutually exclusive. A latent-variable model may impose any combination of them.

EXOGENOUS SELECTION: Many authors assume that u_1 and u_2 are statistically independent conditional on x. It follows that

$$(3.13)\quad P(y|x,z=1) = P[f_1(x)+u_1|x, f_2(x)+u_2\geq 0] = P[f_1(x)+u_1|x] = P(y|x).$$

Thus, independence of u_1 and u_2 conditional on x implies independence of y and z conditional on x, the restriction stated in (3.3). Given (3.3), identification of $P(y|x)$ does not require restrictions on $[f_1(.), f_2(.)]$, but empirical researchers typically impose such restrictions anyway. Most make $f_1(.)$ linear in x.

PARAMETRIC MODELS: A second type of restriction became prominent in the middle 1970s. Suppose that $f_1(.)$ is known up to a finite dimensional parameter β_1, $f_2(.)$ up to a finite dimensional parameter β_2, and the distribution of (u_1, u_2) conditional on x up to a finite dimensional parameter γ. Then

$$(3.14)\quad P(y,z=1|x) = P[f_1(x,\beta_1)+u_1, f_2(x,\beta_2)+u_2\geq 0|x,\gamma].$$

The left-hand-side of (3.14) is identified by the censored sampling process. The right-hand-side is a function of the parameters $(\beta_1, \beta_2, \gamma)$. If there is only one value of $(\beta_1, \beta_2, \gamma)$ solving (3.14), then $P(y|x)$ is identified.

Parametric latent variable models have usually been studied through analysis of $E(y|x,z=1)$. Following the practice in the literature, assume that $E(u_1, u_2|x) = 0$. Then

$$(3.15a)\quad E(y|x) = f_1(x,\beta_1)$$

$$(3.15b)\quad E(y|x,z=1) = f_1(x,\beta_1) + E[u_1|x,f_2(x,\beta_2)+u_2\geq 0,\gamma]$$
$$\equiv f_1(x,\beta_1) + g(x,\beta_2,\gamma).$$

The left-hand-side of (3.15b) is identified by the censored-sampling process. The parameter β_1 is identified, hence $E(y|x)$, if there is only one value of (β_1,β_2,γ) solving (3.15b).

The most widely applied model makes $f_1(.)$ and $f_2(.)$ linear functions, (u_1,u_2) statistically independent of x, and the distribution of (u_1,u_2) normal with mean zero and unrestricted correlation; the variance of u_1 is unrestricted but that of u_2 is set equal to one as a normalization. In this case,

$$(3.16a)\quad E(y|x) = x'\beta_1$$

$$(3.16b)\quad E(y|x,z=1) = x'\beta_1 + \gamma\phi(x'\beta_2)/\Phi(x'\beta_2),$$

where $\phi(.)$ and $\Phi(.)$ are the standard normal density and distribution functions and where $\gamma = E(u_1u_2)$. Identification of β_1 hinges on the fact that the linear function $x'\beta_1$ and the nonlinear function $\gamma\phi(x'\beta_2)/\Phi(x'\beta_2)$ affect $E(y|x,z=1)$ in different ways.

There is a common perception that the normal-linear model generalizes the model assuming exogenous selection. In fact, the two models are not nested. The normal-linear model permits u_1 and u_2 to be dependent but assumes linearity of $[f_1(.),f_2(.)]$, normality of (u_1,u_2), and independence of (u_1,u_2) from x. The exogenous-selection model assumes u_1 and u_2 to be independent

conditional on x but does not restrict $P(u_1|x)$ or $P(u_2|x)$. Nor does it restrict the form of $[f_1(.),f_2(.)]$.

INDEX MODELS: By the early 1980s, parametric models were increasingly criticized. Several articles reported that estimates of $E(y|x)$ obtained under the normal-linear model are sensitive to misspecification of the distribution of (u_1,u_2) conditional on x. Hurd (1979) showed the consequences of heteroskedasticity; Arabmazar and Schmidt (1982) and Goldberger (1983) described the effect of non-normality. Concern with this led to the development of a third type of latent-variable model.

Let h(x) be a known index; that is, a many-to-one function of x. Assume that $f_2(x)$ and the distribution of (u_1,u_2) vary with x only through h(x). Then

$$(3.17a) \quad E(y|x) = f_1(x)$$

$$(3.17b) \quad E(y|x,z=1) = f_1(x) + E\{u_1|x,f_2[h(x)]+u_2 \geq 0\}$$
$$\equiv f_1(x) + g[h(x)].$$

Let (x_0,x_1) denote a pair of points in the support of x such that $h(x_0) = h(x_1)$. For each such pair, (3.17) implies that

$$(3.18) \quad E(y|x=x_0,z=1) - E(y|x=x_1,z=1) = E(y|x=x_0) - E(y|x=x_1).$$

The left-hand side of (3.18) is identified by the sampling process; hence the contrast $E(y|x=x_0) - E(y|x=x_1)$ is identified.

The usefulness of this result depends on the size of the sets $[(x_0, x_1): h(x_0) = h(x_1)]$. The index assumption with the greatest identifying power is that in which h(.) is constant on X. Then (3.18) identifies $E(y|x)$ up to an additive constant. At the other extreme is the trivial case in which h(.) is one-to-one. Then $h(x_0) = h(x_1)$ if and only if $x_0 = x_1$, so (3.18) is uninformative.

The practice has been to combine an index assumption with other restrictions. Robinson (1988) combines an index assumption with the assumption that $f_1(.)$ is linear. In Ahn and Powell (1992) and Cosslett (1991), the index h(.) is not a priori known but assumptions are imposed that make h(.) nonparametrically estimable from the available data on (z,x).

3.2.2. Statistical Mixture Models

Statisticians analyzing censored data often assume that selection is ignorable; that is, hypothesis (3.3). The term "nonignorable" selection is used to cover all situations in which y and z are dependent conditional on x (see, for example, Rubin, 1987).

Some statisticians advocate direct imposition of restrictions on the censored distribution $P(y|x,z=0)$, an approach called "mixture modelling." Suppose that $P(y|x,z=0)$ is known to be a member of a class Γ_{0x} of probability distributions. Then the restriction of $P(y|x)$ to the set given in (3.2) can be improved to

$$(3.19) \quad P(y|x) \in [P(y|x,z=1)P(z=1|x) + \gamma P(z=0|x),\ \gamma \in \Gamma_{0x}].$$

Rubin (1987, Section 6.2) suggests that one not only express prior information by limiting the censored distribution to a set Γ_{0x} but also that one might place a subjective probability distribution on the elements of Γ_{0x}. This then induces a subjective distribution on the elements of the set (3.19) of possible values of $P(y|x)$. Such Bayesian "sensitivity analysis" is feasible only if the set Γ_{0x} is sufficiently small; otherwise a subjective distribution cannot be placed on Γ_{0x}. The practice has been to make Γ_{0x} a finite set or at most a finite dimensional set of distributions. The case of no prior information, in which Γ_{0x} is the set of all distributions, has not received attention in the statistics literature.

TWO WORLD VIEWS: Econometric latent-variable models and statistical mixture models express different ideas about the nature of the selection problem and imply different conclusions about the appropriate way to assert prior information. From the latent-variable-model perspective, the censored distribution is a derived quantity, not a primitive concept. Hence, a researcher who thinks in latent-variable terms finds it difficult to judge the plausibility of restrictions imposed on $P(y|x,z=0)$. From the mixture-model perspective, $P(y|x,z=0)$ is a primitive so it is natural to assert prior information through restrictions on this distribution. Mixture modellers find it difficult to interpret prior information stated as restrictions on latent variable models. The different world views expressed in latent-variable and mixture models have been aired in Wainer (1986,1989).

The conflicting econometric and statistical perspectives on the selection problem recalls a closely related conflict regarding the analysis of discrete data. Econometricians have typically asserted prior information through latent-variable models of discrete choice. Many statisticians have imposed restrictions through the mixture model, referred to as discriminant analysis in that context. See Manski (1981) or Manski and McFadden (1981) for a discussion and references.

3.2.3. Exclusion Restrictions

Empirical analyses often assume that some component of the regressor vector x has no effect on the outcome y but does affect whether y is observed. To formalize this idea, we let $x \equiv (w,v)$ and assume that, holding w fixed, $P(y|w,v)$ does not vary with v but $P(z|w,v)$ does vary with v. The regressor component v is variously said to be an "instrumental variable" or to satisfy an "exclusion restriction."

It has long been recognized that an exclusion restriction may have identifying power when bundled with other assumptions; an example will be given in Section 3.3.1. It is also of interest to determine the identifying power of an exclusion restriction alone, in the absence of other information. This question has been addressed in Manski (1990a,1992a).

The simple result is that an exclusion restriction allows one to replace the bounds available in the absence of prior information with the intersection of these bounds across all values of v. To

see this, let f(w,v) denote the feature of P(y|w,v) that is of interest, perhaps the conditional distribution function or the conditional median. Given an exclusion restriction, f(w,v) must remain constant as w is held fixed and v is varied. Hence, f(w,v) must lie within all of the no-prior-information bounds holding at the different values of v.

For example, the bound (3.6) on the mean of a bounded function of y is replaced by the tighter bound

$$(3.20)\quad \sup_{v} E[g(y)z+K_{0g}(1-z)|w,v] \leq E[g(y)|w] \leq \inf_{v} E[g(y)z+K_{1g}(1-z)|w,v]$$

and the bound (3.11) on the conditional α-quantile is replaced by

$$(3.21)\quad \sup_{v} r[\alpha,(w,v)] \leq q(\alpha,w) \leq \inf_{v} s[\alpha,(w,v)].$$

These new bounds improve on those given in Section 3.1 if the exclusion restriction is non-trivial, in the sense that P(z|w,v) does vary with v. The new bounds do not, however, identify E[g(y)|w] or q(α,w). To achieve point identification generally requires prior information stronger than an exclusion restriction.

3.3. SWITCHING REGRESSIONS

In the selection problem, y is observed when z = 1 and no outcome is observed when z = 0. The literature on switching regressions considers the somewhat richer sampling process in which y is observed when z = 1 and another outcome, say s, is observed when z = 0. The quantities identified by the switching regression sampling process are $P(z|x)$, $P(y|x,z=1)$, and $P(s|x,z=0)$.

In the absence of prior information, observation of s reveals nothing about y. Hence the selection and switching regression problems are equivalent from the perspective of identification of $P(y|x)$. Given prior information, observation of s may be informative regarding $P(y|x)$, as is shown below.

3.3.1. Shifted Outcomes

A rather strong form of prior information that has been applied frequently assumes that there exists a constant $\tilde{N}$ such that

$$(3.22) \quad P(y=s+\nu|x) = 1, \quad \text{all } x.$$

Thus, y and s are assumed to differ by a constant, so that y is a shifted version of s. The implications of (3.22) have been studied by Heckman (1978), Heckman and Robb (1985) and Robinson (1989). Their findings are paraphrased here.

It follows from (3.22) that for all t,

$$(3.23)\quad P(y\le t|x,z=0) = P(s\le t-\nu|x,z=0).$$

Hence,

$$(3.24)\quad P(y\le t|x) = P(y\le t|x,z=1)P(z=1|x) + P(y\le t|x,z=0)P(z=0|x)$$
$$= P(y\le t|x,z=1)P(z=1|x) + P(s\le t-\nu|x,z=0)P(z=0|x).$$

Thus $P(y|x)$ is known up to a family of distributions indexed by the shift parameter ν.

The parameter ν is identified if $E(y|x)$ satisfies an exclusion restriction. It follows from (3.22) that

$$(3.25)\quad E(y|x) = E(y|x,z=1)P(z=1|x) + E(s|x,z=0)P(z=0|x) + \nu P(z=0|x).$$

Let $x = (w,v)$ and suppose it is known that $E(y|w,v=v_0) = E(y|w,v=v_1)$, where v_0 and v_1 are distinct values of v. Then (3.25) implies that

$$(3.26)\quad \nu[P(z=0|w,v=v_1)-P(z=0|w,v=v_0)] =$$
$$E(y|w,v=v_0,z=1)P(z=1|w,v=v_0) + E(s|w,v=v_0,z=0)P(z=0|w,v=v_0)$$
$$-E(y|w,v=v_1,z=1)P(z=1|w,v=v_1) - E(s|w,v=v_1,z=0)P(z=0|w,v=v_1).$$

Hence ν is identified provided that $P(z=0|w,v=v_1) \neq P(z=0|w,v=v_0)$.

3.3.2. Ordered Outcomes

The combination of shifted outcomes and an exclusion restriction yields great identifying power, but requires strong prior information. It is also of interest to learn what can be accomplished with other, perhaps more plausible assumptions. In Manski (1992a), I consider the case in which it is known that

$$(3.27) \quad P(y \geq s \mid x) = 1.$$

This "ordered outcomes" assumption may be warranted in some analyses of medical and other treatments. For example, suppose that a cancer patient is treated by chemotherapy (z=1) or by placebo (z=0). Let the outcomes y and s be life-span following each treatment. Then it may be warranted to assume that $y \geq s$ for all patients.

It follows from (3.27) that for all t,

$$(3.28) \quad P(y \leq t \mid x, z=0) \leq P(s \leq t \mid x, z=0).$$

Hence,

$$\begin{aligned} (3.29) \quad P(y \leq t \mid x) &= P(y \leq t \mid x, z=1)P(z=1 \mid x) + P(y \leq t \mid x, z=0)P(z=0 \mid x) \\ &\leq P(y \leq t \mid x, z=1)P(z=1 \mid x) + P(s \leq t \mid x, z=0)P(z=0 \mid x). \end{aligned}$$

The upper bound on $P(y \leq . \mid x)$ given in (3.29) improves on the bound available if s were not observed when z = 0.

3.3.3. Selection by the Ordering of Outcomes

I also consider the class of problems in which one observes either the smaller or the larger of y and s. Suppose first that one observes the smaller of y and s, so that

(3.30) $z = 1[y \leq s]$.

Examples include the short-side model of markets in disequilibrium (see Maddala, 1983) and the competing-risks model of survival analysis (see Kalbfleisch and Prentice, 1980).

If (3.30) holds, then $z = 0 \iff y > s$. Hence, for all t,

(3.31) $P(y \leq t | x, z=0) \leq P(s \leq t | x, z=0)$.

This is the same finding as was reported in (3.28) under the assumption that outcomes are ordered. So here, as there, (3.29) is an upper bound on $P(y \leq t | x)$.

Now consider the case in which one observes the larger of y and s, so that

(3.32) $z = 1[y \geq s]$.

Examples of this switching rule include economic models of schooling and occupational choice in which decision makers select the alternative yielding the higher income. Here $z = 0 \iff y < s$ so

(3.33) $P(y \leq t | x, z=0) \geq P(s \leq t | x, z=0)$

and

(3.34) $P(y \leq t | x) \geq P(y \leq t | x, z=1) P(z=1 | x) + P(s \leq t | x, z=0) P(z=0 | x)$.

This lower bound on $P(y \leq . | x)$ improves on that available if s were not observed when $z = 0$.

3.4. IDENTIFICATION OF TREATMENT EFFECTS

In the classical formalization of treatment effects, there are two mutually exclusive treatments, labelled 0 and 1. Each member of the population is characterized by values for the variables (y,s,z,x). Variable y is the outcome that would be observed if a person were to receive treatment 1 and s is the outcome that would be observed if the person were to receive treatment 0. Of these two outcomes, one is realized and the other is latent; y is realized if $z = 1$ and s is realized if $z = 0$.

This sampling process is the same as in the switching regression problem. The analysis of treatment effects differs from that of switching regressions only in that there is a different object of interest. The researcher is not concerned with the conditional distribution $P(y|x)$ per se but rather with the treatment effect

(3.35) $T(x) \equiv E(y-s|x) = E(y|x) - E(s|x)$.

Defined in this way, the treatment effect measures the change in average outcome if one were to replace a hypothetical situation in which a person with attributes x were exogenously assigned to treatment 0 with another hypothetical situation in which a person with attributes x were exogenously assigned to treatment 1.[6]

The identification analysis of Sections 3.1 through 3.3 applies directly to the treatment effect (3.35). Section 3.4.1 examines identification in the absence of prior information. Section 3.4.2 shows the identifying power of various forms of prior information and, in a cautionary vein, illustrates the flawed conclusions that can result from the imposition of incorrect assumptions. Section 3.4.3 briefly discusses social experimentation as an approach to securing identifying information.

3.4.1. In the Absence of Prior Information

If either y or s is unbounded, then the sampling process alone reveals nothing about the classical treatment effect. On the other hand, the sampling process alone bounds T(x) if y and s are both bounded random variables. In particular, suppose that y and s both lie in the interval $[K_0, K_1]$. Then it follows immediately from (3.6) that

$$(3.36)\quad T(x) \in$$
$$[K_0 P(z=0|x)+E(y|x,z=1)P(z=1|x)-E(s|x,z=0)P(z=0|x)-K_1 P(z=1|x),$$
$$K_1 P(z=0|x)+E(y|x,z=1)P(z=1|x)-E(s|x,z=0)P(z=0|x)-K_0 P(z=1|x)].$$

The lower bound on $T(x)$ is the difference between the lower bound on $E(y|x)$ and the upper bound on $E(s|x)$. The upper bound on $T(x)$ is determined similarly.

The width of the bound (3.36) is K_1-K_0. If no data were available, $T(x)$ could lie anywhere in the interval $[K_0-K_1, K_1-K_0]$. Thus, the sampling process alone allows one to restrict the treatment effect to one-half of its logically possible range. Observe that the sampling process does not identify the sign of the treatment effect; the bound (3.36) necessarily covers zero.

The case in which y and s are binary outcomes is of particular practical interest. In many applications the treatment outcome is a logical indicator taking the value one or zero. For example, the outcome of a medical treatment may be (cured = 1, not cured = 0); the outcome of a job training program may be (employed = 1, not employed = 0). In such cases, $E(y|x) = P(y=1|x)$, $E(s|x) = P(s=1|x)$, $K_0 = 0$, $K_1 = 1$, and the bound (3.36) becomes

$$(3.37)\quad T(x) \in [P(y=1|x,z=1)P(z=1|x)-P(s=1|x,z=0)P(z=0|x)-P(z=1|x),\ P(z=0|x)+P(y=1|x,z=1)P(z=1|x)-P(s=1|x,z=0)P(z=0|x)].$$

3.4.2. With Prior Information

The bound of the preceding section can be improved if prior information is available. For example, an exclusion restriction may be available or it may be known that subjects are always assigned

the better of the two treatments. In these cases, the same argument used to derive (3.36) can be used to obtain tighter bounds.

Given enough prior information, the treatment effect is identified. Suppose that assignment to treatment is exogenous. Then

$$(3.38)\quad E(y|x) - E(s|x) = E(y|x,z=1) - E(s|x,z=0).$$

Or suppose that the shifted-outcomes hypothesis (3.22) holds. Then the treatment effect is the constant ν for all values of x. As was shown in Section 3.3, ν is identified if an exclusion restriction is available.

Another route to identification is to invoke a latent-variable model with enough structure to identify $E(y|x)$ and $E(s|x)$. Parametric models suffice, but the index models studied in recent years do not identify treatment effects. It was shown in Section 3.2.1 that these models at most identify $E(y|x)$ and $E(s|x)$ up to additive constants. So $E(y|x) - E(s|x)$ is not identified.

The obvious issue arising with the use of prior information to identify treatment effects is that different assumptions may yield different conclusions. I shall give two examples of how a researcher assuming exogenous selection into treatment can reach incorrect conclusions if assignment to treatment is not actually exogenous. These examples are highly relevant because exogenous selection is commonly assumed in practice (see the discussion of regression contrasts following the examples). Moreover, as we observed in Section 3.1, the hypothesis of exogenous selection

cannot be refuted empirically in the absence of prior information. This means that if a researcher believes that selection is exogenous, no data can persuade him that he is wrong.

EXAMPLE 1: Suppose that outcomes are shifted, with $y = s+\nu$ and $\nu > 0$; hence $T(x)$ is the same positive value ν for all x. Suppose that treatments are assigned based on the magnitude of y; for some real constant k, $z = 1$ if $y < k$ and $z = 0$ if $y \geq k$. Then $E(y|x,z=1) = E(y|x,y<k) < k$ and $E(s|x,z=0) = E(s|x,s\geq k-\nu) \geq k-\nu$. Hence, $E(y|x,z=1) - E(s|x,z=0) < \nu$.

Suppose that $Prob(k-\nu\leq s<k|x) = 0$. Then $E(s|x,s\geq k-\nu) = E(s|x,s\geq k) \geq k$ so $E(y|x,z=1) - E(s|x,z=0) < 0$. In this case, a researcher who believes that assignment to treatment is exogenous would improperly conclude not only that the treatment effect is less than ν but that it is negative.

EXAMPLE 2: Suppose that $s = 0$ for all members of the population and that y is a binary variable taking the value - or 1 with probabilities 1-p and p for all x; thus, $T(x) = -\infty$ as long as $p < 1$. Suppose that each member of the population selects the better of the two treatments; thus, $z = 1$ if $y = 1$ and $z = 0$ if $y = -\infty$. Then $E(y|x,z=1) - E(s|x,z=0) = 1 - 0 = 1$. So a researcher who believes that assignment to treatment is exogenous would improperly conclude that the treatment effect is 1 when it actually is $-\infty$.

REGRESSION CONTRASTS AS TREATMENT EFFECTS: The widespread practice of interpreting regression contrasts as treatment effects rests on the assumption that selection into treatment is exogenous. Suppose that, given a random sample of observations of (y,x), one estimates the regression $E(y|x)$, computes a contrast $E(y|x=x_1) - E(y|x=x_0)$ for specified x_0 and x_1, and interprets the contrast as the expected change in y if a person with attributes x_0 were to be given attributes x_1 instead. This interpretation requires an exogenous-selection assumption.

To see why, let us recast the problem in the language of the treatment-effects literature by assuming that each member of the population is characterized by values for the variables $[y(x),z(x),x\epsilon X]$. Here $z(x)$ is an indicator function taking the value 1 at a person's actual regressor value and 0 at all other of the logically possible regressor values X. Variable $y(x)$ is the outcome that would be observed if a person were to be assigned regressor value x. Of these outcomes, $y(x)$ is realized if and only if $z(x) = 1$. Thus, the function $y(.)$ is latent at all regressor values except the one that a person actually experiences. The realized outcome is

$$(3.39) \quad y = \sum_{x\epsilon X} y(x)z(x).$$

This setup is the same as that of the classical treatment effect problem except that now there are more than two treatments; each value of x defines a different treatment.

With this background, we may define the treatment effect $E[y(x_1)] - E[y(x_0)]$ to be the change in average outcome that would be observed if one were to replace a hypothetical situation in which a person were exogenously assigned regressor value x_0 with another hypothetical situation in which that person were exogenously assigned regressor value x_1. In general,

$$(3.40)\quad E(y|x=x_1) - E(y|x=x_0) = E[y(x_1)|z(x_1)=1] - E[y(x_0)|z(x_0)=1]$$
$$\neq E[y(x_1)] - E[y(x_0)].$$

But the second equality does hold if the random outcome function $y(.)$ is statistically independent of the random treatment-selection function $z(.)$.

3.4.3. Social Experimentation

Recognizing that flawed assumptions may yield flawed conclusions, some social scientists advocate that researchers take control of the sampling process by performing social experiments, with subjects randomly assigned to different treatments. In principle, randomization yields exogenous selection of treatments, so a researcher analyzing experimental data can feel confident that the assumption of exogenous selection is valid.

Discussion of social experimentation has at times been highly contentious. In the mid-1980s, various advocates of experimentation asserted that, as a consequence of the selection problem, no reliable inferences can be made from empirical analysis of actual

population outcomes. They recommended that efforts to analyze naturally occurring outcomes be abandoned (see Bassi and Ashenfelter, 1986; Lalonde, 1986; and Coyle et al., 1989). This position has since been embraced by some policymakers. For example, in a recently published letter to the General Accounting Office of The U.S. Congress, an Assistant Secretary of the U.S. Department of Health and Human Services wrote this about the evaluation of training programs for welfare recipients:

> In fact, nonexperimental research of training programs has shown such methods to be so unreliable, that Congress and the Administration have both insisted on experimental designs for the Job Training Partnership Act (JTPA) and the Job Opportunities and Basic Skill (JOBS) programs.

(Barnhart, 1992).

Calls for exclusive reliance on experimentation are unwarranted. Focussing on the identification problems that arise in the analysis of actual population outcomes, proponents of social experiments have ignored the substantial difficulties that arise in executing experiments of interest and in extrapolating from experiments to settings of practical concern. For discussion of these problems, see Manski and Garfinkel (1992).

4. Identification of Endogenous Social Effects From Outcome Data

The broad idea that individuals are influenced by their social environments covers a wide variety of distinct phenomena, from the anonymous process by which markets determine prices to the intensely personal interactions occurring within families. An important objective of social science research is to learn the channels through which society affects the individual. But progress has been limited. Competing hypotheses abound and empirical analysis seems unable to distinguish among them.

Perhaps most notorious is the longstanding interdisciplinary split between economists and other social scientists. Whereas sociologists and social psychologists hypothesize that society affects individuals in myriad ways, economists often assume that society acts on individuals only by constraining their opportunities. Many economists regard such central sociological concepts as norms and reference groups as spurious phenomena explainable by processes operating entirely at the level of the individual. (See, for example, the Friedman, 1957, criticism of Duesenberry, 1949).[7]

Leaving economists aside and restricting attention to sociologists, one still does not find consensus on the nature of social effects. Consider the ongoing debate about the meaningfulness of the concept of the underclass and the related controversy about the existence and nature of neighborhood effects. Or consider the split between those sociologists who take class, ethnic group, or gender as the fundamental unit of analysis, and those who view

society as a collection of heterogeneous individuals, families, or households.

Why do such different perspectives on the nature of social effects persist? Why do we not converge to common conclusions? The core problem is that outcome data, which have been our main source of empirical evidence, can reveal the structure of social effects only if combined with substantial prior information.

Social scientists have long been aware of some aspects of the problem. For over fifty years, economists have studied the conditions under which observations of market-determined prices and quantities reveal the demand behavior of consumers and the supply behavior of firms (see, for example, Fisher, 1966). Over twenty years ago, sociologists were sensitized to the problem of distinguishing contextual effects from correlated individual effects; Hauser (1970) offers an informative and entertaining case study.

Nevertheless, the problem of identifying social effects from outcome data has many poorly understood aspects. In recent work, I have sought to add to our knowledge by analyzing the identifiability of a class of "endogenous" social effects (Manski, 1991b). I summarize and elaborate on this work here.

Section 4.1 introduces the question of interest informally. Section 4.2 uses a simple linear model to examine the identification of endogenous effects relative to contextual effects and non-social phenomena. Section 4.3 briefly considers some more general models. Section 4.4 calls attention to the critical need for reference-group information to identify social effects.

4.1. ENDOGENOUS, CONTEXTUAL, ECOLOGICAL, AND CORRELATED EFFECTS

Consider the following four distinct phenomena, the first two of which are social in nature and the second two non-social:

endogenous effects, wherein the propensity of an individual to behave in some way varies with the prevalence of that behavior in some reference group.[8]

contextual effects, wherein the propensity of an individual to behave in some way varies with the distribution of background characteristics in the reference group.[9]

ecological effects, wherein individuals in the same reference group tend to behave similarly because they face similar institutional environments.

correlated individual effects, wherein individuals in the same reference group tend to behave similarly because they have similar individual characteristics.

An example may help to clarify the distinction. Consider the high school achievement of a teenage youth. There is an endogenous effect if, all else equal, individual achievement varies with the average achievement of the students in the youth's high school, ethnic group, or other reference group. There is a contextual

effect if achievement varies with, say, the socioeconomic composition of the reference group. There is an ecological effect if youth in the same school or other reference group tend to achieve similarly because they receive similar instruction. There is a correlated individual effect if youth in the same school tend to have similar family backgrounds and these background characteristics affect achievement.

The question of interest is whether the two types of social effects can be distinguished from one another and from the non-social effects. This question is of practical importance because the different effects have distinct policy implications. Consider, for example, an educational intervention providing tutoring to some of the students in a school but not to the others. If individual achievement increases with the average achievement of the students in the school, then an effective tutoring program not only directly helps the tutored students but, as their achievement rises, indirectly helps all students in the school, with a feedback to further achievement gains by the tutored students. Contextual effects do not generate this "social multiplier."

Although endogenous and contextual effects differ conceptually and in their policy implications, these two types of social effect have often been confused. For example, studies of school integration, typified by Coleman et al. (1966), seem to have in mind an endogenous social effect, wherein the achievement of each student is affected by the mean achievement of the students in the same school. But these studies generally estimate contextual-effects

models, wherein the achievement of each student is affected by the racial composition of his school.

The same tension appears in recent analyses of neighborhood effects. The theoretical section of Crane (1991) poses an "epidemic" model of endogenous neighborhood effects, wherein a teenager's school dropout and childbearing behavior is influenced by the neighborhood frequency of dropout and childbearing. But Crane estimates a contextual-effects model, wherein a teenager's behavior depends on the occupational composition of her neighborhood. This juxtaposition of endogenous-effect theorizing and contextual-effect empirical analysis also appears in Jencks and Mayer (1989).

4.2. IDENTIFICATION OF A LINEAR MODEL: THE REFLECTION PROBLEM

4.2.1. Model Specification

Consideration of a relatively simple linear model suffices to explain the problems that arise in identifying social effects.

Let each member of the population be characterized by a value for (y,x,z,u). Here y is a scalar outcome (e.g. a youth's achievement in high school), x are attributes characterizing an individual's reference group (e.g. a set of dummy variables denoting a youth's high school and/or ethnic group), and (z,u) are attributes that directly affect y. A researcher observes a random sample of realizations of (y,x,z). Realizations of u are not observed.

I shall assume that

$$(4.1)\quad y = \alpha + \beta E(y|x) + E(z|x)'\gamma + x'\delta_1 + z'\eta + u, \quad E(u|x,z) = x'\delta_2,$$

where $(\alpha,\beta,\gamma,\delta_1,\delta_2,\eta)$ is a parameter vector. Model (4.1) implies that the mean regression of y on (x,z) has the linear form

$$(4.2)\quad E(y|x,z) = \alpha + \beta E(y|x) + E(z|x)'\gamma + x'(\delta_1+\delta_2) + z'\eta.$$

Empirical studies of social effects have generally assumed that values of the regressors $[E(y|x),E(z|x),x,z]$ are assigned exogenously to individuals. Hence the parameters $(\alpha,\beta,\gamma,\delta_1,\delta_2,\eta)$ are the treatment effects associated with a unit change in each regressor, holding the others fixed (see Section 3.4.2).[10]

If $\beta \neq 0$, the linear regression (4.2) expresses an endogenous social effect: a person's response y varies with $E(y|x)$, the mean of the endogenous variable y among those persons in the reference group described by x.[11] If $\gamma \neq 0$, the model expresses a contextual effect: y varies with $E(z|x)$, the mean of the exogenous variables z among those persons in the reference group. If $\delta_1 \neq 0$, the model expresses an ecological effect: y varies directly with x. If $\delta_2 \neq 0$, the model expresses correlated individual effects: persons in the reference group x tend to have similar unobserved attributes u. The parameter η expresses the direct effect of z on y.

4.2.2. Identification of the Parameters

We are interested in identification of the parameter vector $(\alpha,\beta,\gamma,\delta_1,\delta_2,\eta)$. To focus attention on this question, I shall

assume that either (i) x has discrete support or (ii) y and z have finite variances and the two regressions $E(y|x)$ and $E(z|x)$ appearing as regressors in (4.2) are continuous on the support of x. As indicated in Section 2.1, either of these assumptions implies that the random sampling process identifies $E(y|x)$ and $E(z|x)$ on the support of x. So we can treat the two regressions as known and focus attention on the parameters.[12]

One aspect of the identification problem can be seen immediately by inspection of (4.2). That is, ecological effects cannot be identified relative to correlated individual effects. The sum $(\delta_1+\delta_2)$ may be identified but not δ_1 and δ_2 separately.

Less obvious is the "reflection" problem that arises out of the presence of $E(y|x)$ as a regressor in (4.2). Integrating both sides of (4.2) with respect to z reveals that $E(y|x)$ solves the "social equilibrium" equation

$$(4.3)\quad E(y|x) = \alpha + \beta E(y|x) + E(z|x)'\gamma + x'(\delta_1+\delta_2) + E(z|x)'\eta.$$

Provided that $\beta \neq 1$, equation (4.3) has a unique solution, namely

$$(4.4)\quad E(y|x) = [\alpha+E(z|x)'(\gamma+\eta)+x'(\delta_1+\delta_2)]/(1-\beta).$$

Thus, model (4.2) implies that $E(y|x)$ is a linear function of $[1,E(z|x),x]$, where "1" denotes the constant. It follows that the parameters α, β, γ, $(\delta_1+\delta_2)$ are all unidentified. In particular, endogenous effects cannot be distinguished from contextual effects.

What is identified? Inserting (4.4) into (4.2) we obtain the linear reduced form model

$$(4.5)\quad E(y|x,z) = \alpha/(1-\beta) + E(z|x)'[\gamma/(1-\beta)+\eta\beta/(1-\beta)] + x'[(\delta_1+\delta_2)/(1-\beta)] + z'\eta.$$

The composite parameters $\alpha/(1-\beta)$, $\gamma/(1-\beta)+\eta\beta/(1-\beta)$, $(\delta_1+\delta_2)/(1-\beta)$, and η are identified if the regressors $[1,E(z|x),x,z]$ are linearly independent. Identification of the composite parameters does not enable one to distinguish among the various social and non-social effects but does permit one to test the hypothesis that some social or non-social effect is present. If $\gamma/(1-\beta)+\eta\beta/(1-\beta)$ is non-zero, then either β and γ must be non-zero; so some social effect is present. If $[(\delta_1+\delta_2)/(1-\beta)]$ is non-zero, then either δ_1 and δ_2 must be non-zero; so some non-social effect is present.

Even these relatively weak identification findings are tenuous. The required linear independence of $[1,E(z|x),x,z]$ is a non-trivial condition that can fail in various ways, including the following:

(a) The attributes x defining reference groups may be a subset of the attributes z directly affecting outcomes. Suppose that $z = (x,w)$ for some vector w. Then $[1,E(z|x),x,z] = [1,\{x,E(w|x)\},x,\{x,w\}]$ is linearly dependent through the appearance of x in three locations.

(b) The attributes z directly affecting outcomes may be a subset of the attributes x defining reference groups. Suppose that $x = (z,w)$ for some vector w. Then $[1,E(z|x),x,z] =$

[1,z,(z,w),z] is linearly dependent through the appearance of z in three locations.

(c) The attributes z directly affecting outcomes may be mean-independent of the attributes x defining reference groups. Suppose that $E(z|x) = z_0$, for some constant vector z_0. Then $[1,E(z|x),x,z] = [1,z_0,x,z]$ is linearly dependent through the appearance of the two constants 1 and z_0.

(d) The regression $E(z|x)$ may be a linear function of x as, for example, occurs if (z,x) are distributed multivariate normal. Suppose that $E(z|x) = Ax$ for some parameter matrix A. Then $[1,E(z|x),x,z] = [1,Ax,x,z]$ is linearly dependent through the appearance of Ax and x.

Taken together, these conditions say that identification fails unless the attributes z and x are "moderately" related in a nonlinear manner. They must be neither functionally dependent (conditions a and b), mean independent (condition c), nor linearly mean-dependent (condition d).

4.2.3. Parameter Restrictions

The possibilities for identification improve if one has prior information restricting some of the parameters. The most common restrictions are assumptions that some parameter values are zero, so that the corresponding effect is null. Suppose it is known that $\beta = 0$, so that there is no endogenous effect. Then the contextual-effect parameter γ is identified if $[1,E(z|x),x,z]$ are linearly independent. Or suppose it is known that $\gamma = 0$, so that there is

no contextual effect. Then the endogenous-effect parameter β is identified if $[1,E(z|x),x,z]$ are linearly independent and $\eta \neq 0$.

The only way to relax the linear independence condition is to impose further parameter restrictions. Empirical studies of contextual effects generally assume that $\beta = \delta_1 = \delta_2 = 0$; then γ is identified if $[1,E(z|x),z]$ are linearly independent. Empirical studies of endogenous effects generally assume that $\gamma = \delta_1 = \delta_2 = 0$; then β is identified if $[1,E(z|x),z]$ are linearly independent and $\eta \neq 0$.

4.2.4. Sample Inference

Our primary concern is with identification of the model (4.2), but a discussion of sample inference is warranted.

Empirical studies of contextual effects have always applied a two-stage method to estimate (γ,η). In the first stage, one uses the sample data on (z,x) to estimate $E(z|x)$ nonparametrically; generally x is discrete and the estimate of $E(z|x)$ is a cell-average of the form given in equation (2.1). In the second stage, one estimates (γ,η) by finding the least squares fit of y to $[E_N(z|x),z]$, where $E_N(z|x)$ is the first-stage estimate of $E(z|x)$.

Empirical studies of endogenous effects have also applied a two-stage method to estimate (β,η), but in the guise of a "spatial autocorrelation model. Spatial correlation models have the form

$$(4.6) \quad y_i = \beta W_{iN} Y + z_i'\eta + u_i, \qquad i = 1,\ldots,N.$$

Here $Y = (y_i, i=1,...,N)$ is the $N \times 1$ vector of sample realizations of y and W_{iN} is a specified $1 \times N$ weighting vector; that is, the components of W_{iN} are non-negative and sum to one. The disturbances u are assumed to be normally distributed, independent of x, and the model is estimated by maximum likelihood. See, for example, Cliff and Ord (1981) or Case (1991).

Equation (4.6) states that the behavior of each person in the sample varies with a weighted average of the behaviors of the other sample members. Thus, the spatial correlation model assumes that a social effect is generated within the researcher's sample rather than within the population from which the sample was drawn. This makes sense in studies of small-group interactions, where the sample is composed of clusters of friends, co-workers, or household members; see, for example, Duncan, Haller, and Portes (1968). But it does not make sense in studies of neighborhood and other large-group social effects, where the sample members are randomly chosen individuals. Taken at face value, equation (4.6) implies that the sample members know who each other are and choose their outcomes only after having been selected into the sample.

The spatial correlation model does make sense in studies of large-group interactions if interpreted as a two-stage method for estimating model (4.2). In the first stage, one uses the sample data on (y,x) to estimate $E(y|x)$ nonparametrically, and in the second stage, one estimates (β, η) by finding the least squares fit of y to $[E_N(y|x), z]$, where $E_N(y|x)$ is the first-stage estimate of $E(y|x)$. Many nonparametric estimates of $E(y|x_i)$, including the

local average (2.2), are weighted averages of the form $E_N(y|x_i) = W_{iN}Y$, with W_{in} determining the specific estimate. Hence, estimates of (β,γ) reported in the spatial correlation literature can be interpreted as estimates of (4.2).

THE SAMPLING DISTRIBUTION OF TWO-STAGE ESTIMATES: It is necessary to point out that empirical studies reporting two-stage estimates of social-effects models have routinely misreported the sampling distribution of their estimates. The practice in two-stage estimation of contextual-effects models has been to treat the first-stage estimate $E_N(z|x)$ as if it were $E(z|x)$ rather than an estimate thereof. The literature on spatial correlation models has presumed that equation (4.6) holds as stated and has not specified how the weights W_{iN} should change with N.

Two-stage estimation of social-effects models is similar to other semiparametric two-stage estimation problems whose asymptotic properties have been studied recently. Ahn and Manski (1992), Ichimura and Lee (1991), and others have analyzed the asymptotic behavior of various estimators whose first stage is nonparametric regression and whose second stage is parametric estimation conditional on the first-stage estimate. It is typically found that the second-stage estimate is $\sqrt{N}$-consistent with a limiting normal distribution if the first-stage estimator is chosen appropriately. The variance of the limiting distribution is typically larger than that which would prevail if the first-stage

regression were known rather than estimated. It seems likely that this result holds here as well.

4.3. MORE GENERAL MODELS

The analysis of the preceding section sends a strong warning about the difficulty of inferring social effects from outcome data. Consideration of richer, more realistic models complicates matters further. This section describes several generalizations of model (4.2) and calls attention to the identification issues that they raise.

4.3.1. Nonlinear Models

There is often no good reason to think that social and non-social effects behave linearly. The basic themes of (4.2) are captured by the class of models of the form

$$(4.7) \quad E(y|x,z) = f[E(y|x),E(z|x),x,z],$$

$f(.,.)$ being a member of some family F of functions on $Y\times Z\times X\times Z$. Whereas (4.2) implied that $E(y|x)$ solved the linear social equilibrium equation (4.3), (4.7) implies that $E(y|x)$ solves the possibly nonlinear social equilibrium equation

$$(4.8) \quad E(y|x) = \int f[E(y|x),E(z|x),x,z]dP(z|x),$$

where $P(z|x)$ is the probability distribution of z conditional on x.

The model (4.7) is coherent if equation (4.8) has a solution. If there is no solution, then the model is internally inconsistent. Sometimes, as when f(.) is linear, there is a unique solution to (4.8). In these situations, $E(y|x)$ can be expressed as a function of $\{P(z|x),x\}$. Then the model (4.7) has a reduced form

$$(4.9) \quad E(y|x,z) = f[g\{P(z|x),x\},E(z|x),x,z],$$

where g(.) gives $E(y|x)$ as a function of $\{P(z|x),x\}$.

In Manski (1992b), I have studied identification of endogenous-effects models with unique social equilibria. Let it be known that

$$(4.10) \quad E(y|x,z) = f[E(y|x),z]$$

and that there is a unique social equilibrium. It is shown that if the form of the function f(.) is not known, then one can hope to identify how f(.) varies with $E(y|x)$ only if z and x are moderately related. In particular, identification is not possible if z and x are either functionally dependent or statistically independent. This result is a nonparametric extension of the linear-independence condition studied in Section 4.2.2.

BINARY RESPONSE MODELS: Perhaps the most common non-linear social effects models in the literature are binary response models. Let y be a binary random variable, so that $E(y|x,z) = P(y=1|x,z)$ and

$E(y|x) = P(y=1|x)$. Assume that, for some continuous and strictly increasing distribution function $H(.)$ on the real line,

$$(4.11) \quad P(y=1|x,z) = H[\alpha+\beta P(y=1|x)+E(z|x)'\gamma+x'\delta-z'\eta],$$

where $(\alpha,\beta,\gamma,\delta,\eta)$ are parameters. This model, which includes the logit and probit models as special cases, has a social equilibrium if $P(y=1|x)$ solves the equation

$$(4.12) \quad P(y=1|x) = \int H[\alpha+\beta P(y=1|x)+E(z|x)'\gamma+x'\delta+z'\eta]dP(z|x).$$

It is shown in Manski (1992b) that equation (4.12) always has at least one solution, so the model is coherent; if $\beta \leq 0$, the solution is unique. The conditions under which the parameters $(\alpha,\beta,\gamma,\delta,\eta)$ are identified have not been established.

Models of form (4.11) have been estimated by two-stage methods. One estimates $P(y=1|x)$ nonparametrically and then estimates (β,γ) by maximizing the quasi-likelihood in which $P_N(y=1|x)$ takes the place of $P(y=1|x)$. Examples include Case and Katz (1991) and Gamoran and Mare (1989). A multinomial response model estimated in this manner appears in Manski and Wise (1983), Chapter 6.

4.3.2. More General Social Effects

So far, we have assumed that social effects are transmitted through $E(y|x)$ and $E(z|x)$, but they may be transmitted through other channels as well. Three are mentioned here. Models incorporating these more general effects bring to bear less prior information, so the identification problem inevitably worsens.

Endogenous and contextual effects may be transmitted not only through $E(y|x)$ and $E(z|x)$ but through the entire conditional distributions $P(y|x)$ and $P(z|x)$. These social forces may affect not only $E(y|x,z)$ but the entire conditional distribution $P(y|x,z)$. If so, then we have the following abstract generalization of (4.7):

$$P(y|x,z) = f[P(y|x),P(z|x),x,z]. \quad (4.13)$$

For example, it is sometimes said that the strength of the effect of social norms on individual behavior depends on the dispersion of behavior in the population. The more homogeneous is reference-group behavior, the stronger the norm. This idea can be expressed by models in which individual outcomes vary not only with the mean outcome of the reference group but also with the variance of the reference-group outcomes.

Another direction for generalization is to allow individuals to be influenced by multiple reference groups, giving more weight to the behavior of some groups than to others. Then (4.14) might generalized even further to

$$(4.14)\quad P(y|(x_m,m=1,...M),z) = f[(P(y|x_m),P(z|x_m),x_m,i=1,...,M),z].$$

Here x_m characterizes the m^{th} reference group.

Yet another direction for generalization is to let the outcome y be a vector rather than a scalar. With this done, we can imagine a simultaneous system of endogenous effects, with reference-group outcomes along each dimension affecting individual outcomes along other dimensions.

4.3.3. Dynamic Models

Some authors, including Alessie and Kapteyn (1991) and Borjas (1991), have estimated the following dynamic version of the linear model (4.2):

$$(4.14)\quad E_t(y|x,z) = \alpha + \beta E_{t-1}(y|x) + E_{t-1}(z|x)'\gamma + x_t'(\delta_1+\delta_2) + z_t'\eta,$$

where E_t and E_{t-1} denote expectations taken at periods t and t-1. The idea is that non-social forces act contemporaneously but social forces act on the individual with a lag.

If $\{E(z|x),x,z\}$ are time-invariant and $-1 < \beta < 1$, the dynamic process (4.14) has a unique stable temporal equilibrium of the form (4.3). If one observes the process in temporal equilibrium, the identification analysis in Section 4.2 holds without modification. On the other hand, if one observes the process out of equilibrium, the recursive structure of (4.14) opens new possibilities for

identification. In particular, $E_{t-1}(y|x)$ is not necessarily a linear function of $[1, E_{t-1}(z|x), x_t]$.

One should not, however, conclude that dynamic models solve the problem of identifying social effects. To exploit the recursive structure of (4.14), a researcher must maintain the hypothesis that the transmission of social effects really follows the assumed temporal pattern. But empirical studies typically provide no evidence for any particular timing. Some authors assume that individuals are influenced by the behavior of their contemporaries, some assume a time lag of a few years, while others assume that social effects operate across generations.

4.4. IDENTIFICATION OF REFERENCE GROUPS

So far, we have assumed that the researcher knows individuals' reference groups. There is substantial reason to question this assumption. Researchers rarely offer any empirical evidence on how individuals form reference groups or, for that matter, on whether the reference-group concept is meaningful to them. One study that does attempt to justify its specification is Woittiez and Kapteyn (1991), who use individuals' responses to questions about their "social environments" as evidence on their reference groups.

Researchers do not try to determine whether individuals actually observe the outcomes of their reference-groups. They assume that individuals are influenced by $E(y|x)$ and $E(z|x)$, but offer no evidence that people perceive these quantities correctly.[13]

4.4.1. Tautological Linear Models

If researchers do not know how individuals form reference groups and perceive reference-group outcomes, then it is reasonable to ask whether outcome data can be used to infer these unknowns. In Manski (1991b), I showed that the answer is negative. Suppose one believes that some vector z directly affects individual outcomes but one does not know the vector x that defines reference groups. Then one cannot reject the hypothesis that x is a superset of z and may not be able to reject the hypothesis that x is a subset of z.

Consider first the hypothesis that x is a superset of z; that is, $x = (z,w)$ for some vector w. Then $E(y|x,z) = E(y|x)$. So model (4.2) holds tautologically with $\beta = 1$ and $\alpha = \gamma = \delta_1 = \delta_2 = \eta = 0$. Thus, one cannot reject the hypothesis that x is a superset of z and that mean reference-group behavior dictates individual behavior.

Now consider the alternative hypothesis that x is a subset of z; that is, $z = (x,w)$ for some vector w. Then $E(y|x,z) = E(y|z)$. If $E(y|z)$ is a linear function $z'c$, c being a parameter vector, then the linear model (4.2) holds tautologically with $\eta = c$ and $\alpha = \beta = \delta_1 = \delta_2 = 0$. Thus, one cannot reject the hypothesis that x is a subset of z and that z dictates individual behavior, with no role for endogenous, contextual, ecological, or correlated effects.

For example, consider a researcher studying student achievement. Suppose that the researcher observes each student's ability and ethnicity. If the researcher specifies x to be (ability,ethnicity) and z to be (ability), he will find that the data are consistent with the hypothesis that individuals do condition on (ability,

ethnicity) to form their reference groups, that individual achievement reflects reference-group achievement, and that ability has no direct effect on achievement. If the researcher specifies x to be (ethnicity) and z to be (ability,ethnicity), he will find that the data are consistent with the hypothesis that reference-group achievement does not affect individual achievement.

4.4.2. Experimental and Subjective Data

It clearly is very difficult to draw conclusions about social effects from outcome data alone. If the identification of social effects is so tenuous, then why is there such a widespread perception that society influences individual behavior in many ways? It may be that this common perception is poorly grounded, fed by flawed interpretations of outcome data. But outcome data are not our only source of evidence on social effects. Prevailing views also rest on evidence from controlled experiments and on subjective data, the statements people make about why they behave as they do. See Jones (1984) for a survey of the experimental literature.

Our analysis of identification from outcome data suggests that experimental and subjective data will have to play an important role in future efforts to learn about social effects. At a minimum, subjective data are necessary to identify reference groups, which are a subjective phenomenon. The problem of identifying subjective phenomena is discussed further in Section 5.

5. Identification of Subjective Phenomena: The Use of Intentions Data

5.1. RESEARCH PRACTICES IN ECONOMICS AND SOCIOLOGY

Policy disputes often reflect disagreement about the roles of objective and subjective forces in determining behavior. Suppose, as do many social scientists, that behavior is determined by objective opportunities and by subjective preferences and expectations. Then we may ask politically sensitive questions such as:

> Do young black males have low labor force participation because (a) jobs are unavailable (opportunities), (b) they believe that jobs are unavailable (expectations), or (c) they don't want to work (preferences)?

Social scientists and concerned citizens agree that this question is meaningful and relevant to social policy. We have not, however, been able to reach consensus on the answer.

Distinguishing the objective and subjective determinants of human behavior may be the most challenging identification problem facing social scientists. It is certainly the problem that most clearly separates the social from the natural sciences, where the units of analysis are not thought of as possessing free will. Yet the inherent difficulty of inference on subjective phenomena does not fully explain our lack of knowledge. Research practices in the various social science disciplines also inhibit progress.

For many years, economists have exercised a self-imposed prohibition on the use of subjective data in empirical analysis.[14]

Instead, they have sought to infer subjective phenomena from data on opportunities and choices. This research approach, referred to as "revealed preference analysis," cannot be used to jointly infer expectations and preferences. So economists have typically imposed assumptions on expectations and attempted to infer preferences from observed choices.[15]

In contrast to economists, social psychologists and sociologists routinely collect and analyze subjective data of many kinds. Unfortunately, the prevailing practice has been to pose loosely-worded questions incapable of revealing much about either expectations or preferences. Moreover, researchers typically theorize verbally rather than mathematically. Hence, it can be difficult to determine whether different researchers interpret the terms "preferences" and "expectations" in a common manner.

As I see it, progress in understanding the objective and subjective determinants of behavior requires that the various social sciences break with their conventions. As long as economists continue to rely exclusively on revealed-preference analysis, they have no hope of determining whether they are making empirically valid inferences on preferences or wrong inferences based on incorrect expectations assumptions. As long as sociologists continue to reason verbally rather than mathematically, their empirical analysis will suffer from conceptual ambiguity.

Some of my recent and ongoing work seeks to fuse what I see as the positive aspects of present economic and sociological research practices: the use of formal decision theory by economists and the

exploitation of subjective data by sociologists. One completed paper examines the conditions under which rational decision makers can learn from the experiences of role models (Manski, 1992b). Another, focussing on the analysis of schooling behavior, critiques the conventional economic practice of assuming that youth have specific expectations of the returns to schooling(Manski, 1992c). In ongoing work, I am attempting to elicit from youth their actual expectations of the returns to schooling.

In this section, I describe my recent work offering a "best-case" decision theoretic treatment of stated intentions, a familiar type of subjective data. The analysis, drawn from Manski (1990b), shows that intentions data have often been mis-interpreted. In doing so, it illustrates the importance of interpreting subjective data in a logically coherent manner.

5.2. THE USE OF INTENTIONS DATA TO PREDICT BEHAVIOR

In surveys individuals are routinely asked to predict their future behavior, that is, to state their intentions. The fertility question asked female respondents in the June 1987 Supplement to the Current Population Survey (CPS) is an example:

> Looking ahead, do you expect to have any (more) children?
> Yes No Uncertain
>
> (U.S. Bureau of the Census,1988)

Responses to such fertility-intentions questions have been used to predict fertility for over fifty years; Hendershot and Placek (1981) review the extensive literature. Data on voting intentions have been used to predict American election outcomes since the early 1900s (see Turner and Martin,1984). Surveys of buying intentions have been used to predict consumer purchase behavior since at least the mid 1940s (see Juster,1966). Perhaps the most extensive use of intentions data has been made by social psychologists, some of whom view intentions as a well-defined mental state that causally precedes behavior (see Fishbein and Ajzen, 1975).

A BEST-CASE ANALYSIS: What information do intentions data convey about future behavior? The answer depends on how people respond to intentions questions and on how they actually behave. In Manski (1990b), I studied the relationship between stated intentions and subsequent behavior under the "best-case" hypothesis that individuals have rational expectations and that their responses to intentions questions are best predictors of their behavior. My objective was to place an upper bound on the behavioral information contained in intentions data and to determine whether prevailing approaches to the analysis of intentions data respect the bound.

I found that much of the literature interprets intentions data in ways that are inconsistent with the best-case analysis. Authors have expected too much correspondence between intentions and behavior. Not finding the expected correspondence, they have

improperly concluded that individuals are poor predictors of their future behavior.

5.2.1. The Survey Question and the Best-Case Response

My analysis focusses on the simplest intentions questions, those that call for yes/no predictions of binary outcomes. Suppose that a person is asked to make a point prediction of some binary choice; that is, a yes/no answer is requested. Let i and y be zero-one indicator variables denoting the survey response and future behavior respectively. Thus i = 1 if the person responds "yes" to the intentions question and y = 1 if his behavior turns out to satisfy the property of interest.

To form his response, a person with rational expectations would begin by recognizing that his future behavior will depend in part on conditions known to him at the time of the survey and in part on events that have not yet occurred. Let s denote the information available to the respondent at the time of the survey. Let z denote the events that have not yet occurred but which will affect his future behavior. Thus z represents uncertainty which will be resolved between the time of the survey and the time at which the behavior is determined. The behavior y is a function of the pair (s,z) and so may be written y(s,z).

Let $P_z|s$ denote the objective probability distribution of z conditional on s. Let P(y|s) denote the objective distribution of y conditional on s. The event y = 1 occurs if and only if the realization of z is such that y(s,z) = 1. Hence

$$(5.1) \quad P(y=1|s) = P_z[y(s,z)=1|s].$$

The content of the rational-expectations hypothesis is that, at the time of the survey, the respondent knows $y(s,.)$ and $P_z|s$; hence he knows $P(y=1|s)$. It does not suffice for the respondent to have a subjective distribution for z, from which he derives a subjective distribution for y. Rational expectations assumes knowledge of the actual stochastic process generating z.

The second part of the best-case hypothesis is that the respondent states his best point prediction of his behavior. The best prediction necessarily depends on the losses the respondent associates with the two possible prediction errors (i=0,y=1) and (i=1,y=0). These losses may be influenced by the wording of the intentions question; for example, the respondent may interpret differently questions that ask what he "expects," "intends," or "is likely" to do. Whatever the loss function, however, the intentions response satisfies the condition

$$(5.2) \quad \begin{aligned} i = 1 &\Rightarrow P(y=1|s) \geq \pi \\ i = 0 &\Rightarrow P(y=1|s) \leq \pi , \end{aligned}$$

where the threshold probability π depends on the loss function. Note that $\pi = .5$ if the loss function is symmetric.

5.2.2. Prediction of Behavior Conditional on Intentions

Now consider a researcher who wishes to use intentions data to predict the behavior of some respondent. The researcher observes the intentions response i. Continuing the theme of a best-case analysis, assume the researcher knows that i satisfies (5.2). Moreover, assume that π is the same for all respondents and that the researcher knows what π is.

The researcher may observe only a subset of the information s available to the respondent. Let x denote the observed component of s. Suppose that the researcher wishes to predict the behavior y conditional on the observed variables x and i. Then he would like to learn the probability $P(y=1|x,i)$. Intentions data do not identify $P(y=1|x,i)$. They do, however, imply a bound.

Let $P_s|xi$ denote the probability distribution of s conditional on the observed pair (x,i). It is the case that

$$(5.3) \quad P(y=1|x,i) = \int P(y=1|s)dP_s|xi.$$

It follows directly from this and from (5.2) that

$$(5.4) \quad P(y=1|x,i=0) \leq \pi \leq P(y=1|x,i=1).$$

This bound expresses all the information about behavior contained in the intentions data. Note that the bound varies with i but not with x.

The foregoing implies that familiar path models attempting to explain behavior as a function of intentions are not consistent with the best-case hypothesis. Consider a logit model

$$(5.5)\quad P(y=1|x,i) = \frac{\exp(x\beta+\gamma i)}{1+\exp(x\beta+\gamma i)},$$

where (β,γ) are parameters. This model has the property

$$(5.6)\quad \begin{aligned} x\beta+\gamma i < 0 &\Rightarrow P(y=1|x,i) < 1/2 \\ x\beta+\gamma i = 0 &\Rightarrow P(y=1|x,i) = 1/2 \\ x\beta+\gamma i > 0 &\Rightarrow P(y=1|x,i) > 1/2. \end{aligned}$$

Suppose that $\pi = 1/2$. Then (5.6) is consistent with (5.4) only if (x,β,γ) satisfies the special property $x\beta \le 0 \le x\beta+\gamma$.

The problem is, of course, not specific to the logit model and the case $\pi = 1/2$. It is characteristic of any path model which attempts to explain y as a function of a linear index $x\beta+\gamma i$.

5.2.3. Prediction Not Conditional on Intentions

Often a researcher wants to predict the behavior of a nonsampled member of the population from which the survey respondents were drawn. Intentions data are available only for the sampled individuals. But some background variables x may be observed for the entire population. In this setting, one may want to predict behavior conditional on these x. Then the quantity of interest is $P(y=1|x)$.

The bound (5.4) implies a bound on $P(y=1|x)$. Observe that

$$(5.7)\quad P(y=1|x) \equiv P(y=1|x,i=0)P(i=0|x) + P(y=1|x,i=1)P(i=1|x).$$

It follows from (5.4) and (5.7) that

$$(5.8)\quad \pi P(i=1|x) \leq P(y=1|x) \leq \pi P(i=0|x) + P(i=1|x).$$

This bound is useful in practice because $P(i=1|x)$ can be estimated nonparametrically from the sample data.

Observe that the bound (5.8), unlike (5.4), varies with x. The bound width, which is $\pi P(i=0|x) + (1-\pi)P(i=1|x)$, may take any value between zero and one, depending on the magnitudes of π and $P(i|x)$. Thus, depending on the application, intentions data may yield a tight or a weak bound on $P(y|x)$. If $\pi = 1/2$, the bound width is 1/2, whatever $P(i|x)$ might be.

It has been known for at least twenty-five years that the sharp relationship

$$(5.9)\quad P(i=1|x) = P(y=1|x)$$

need not hold (see Juster,1966, p.665). Nethertheless, some of the literature continues to consider deviations from this equality as "inconsistencies" in need of explanation. For example, Westoff and Ryder (1977) state:

> The question with which we began this work was whether reproductive intentions are useful for prediction. The basic finding was that 40.5 percent intended more, as of the end of

> 1970, and 34.0 percent had more in the subsequent five years . . . . In other words, acceptance of 1970 intentions at face value would have led to a substantial overshooting of the ultimate outcome. (p. 449)

Seeking to explain the observed "overshooting" of births, the authors state:

> one interpretation of our finding would be that the respondents failed to anticipate the extent to which the times would be unpropitious for childbearing, that they made the understandable but frequently invalid assumption that the future would resemble the present--the same kind of forecasting error that demographers have often made. (p. 449)

More recent demographic work maintains the presumption that deviations from (5.9) require explanation. See, for example, Davidson and Beach (1981) and O'Connell and Rogers (1983).

The best-case hypothesis implies that (5.9) should hold in one very special case; that in which future behavior depends only on the information s available at the time of the survey. In this case, the respondent can forecast his future behavior with certainty. So i always equals y.

In the nondegenerate case where future events z partially determine behavior, the best-case hypothesis does not imply (5.9). A simple example makes the point forcefully. Let $\pi = 1/2$ and let $P(y=1|s) = .51$ for all values of s. Then $P(y=1|x) = .51$ but $P(i=1|x) = 1$ for all values of x. This demonstrates that

individual-level differences between intentions and behavior do not, in general, average-out in the aggregate.

5.2.4. Lessons

The use of intentions data to predict behavior has been controversial. At least some of the controversy is rooted in the flawed premise that divergences between intentions and behavior show individuals to be poor predictors of their futures. Divergences may simply reflect the dependence of behavior on events not yet realized at the time of the survey. Divergences will occur even if responses to intentions questions are the best predictions possible given the available information. The lesson is that researchers should not expect too much from yes/no intentions data.

In principle, the yes/no form of intentions question can be improved upon by asking the respondent to give his probability for the behavior in question. Whereas a yes/no question reveals at most the bounds (5.4) and (5.8) on $P(y|x,i)$ and $P(y|x)$, probability elicitation may reveal $P(y|s)$. See Juster (1966) for an interesting empirical study eliciting probabilistic intentions.

Notes

1. The modern statistical literature uses the term "regression" in a much more general sense to refer to any feature of the probability distribution of y conditional on x, for example the conditional median or variance. So we speak of the mean regression, median regression, variance regression, and so on (see Manski, 1991a). The extrapolation problem discussed in this section applies to all of these senses of regression, not just to the familiar mean regression.

2. The literature on nonparametric regression analysis refers to (2.2) as a "uniform kernel" estimate. The reader who wishes further exposition of this and other nonparametric regression methods may turn to Manski (1991a) and to the opening chapters of Hardle (1990).

I should note that the terms "parametric" and "nonparametric" are conventionally used to distinguish those problems in which the regression is known up to a finite-dimensional parameter from those in which the regression is known to be a member of some non-finite-dimensional space of functions of x. For example, the regression might be known to be a continuous function of x, as in condition (a). Use of the term "nonparametric" to mean that the parameter space is a space of functions is an illogical but firmly entrenched semantic convention.

3. The assumption that y is scalar will be used only in a few places. Most of the analysis extends immediately to situations in which y is a vector.

4. To obtain some intuition for this fact, consider the following thought experiment. Let w be a random variable with $\text{Prob}(w \leq t] = 1-\eta$ and $\text{Prob}(w=s) = \eta$, where $s > t$. Suppose w is perturbed by moving the mass at s to some $s_1 > s$. Then $P(w \leq r)$ remains unchanged for $r < s$ and falls by at most η for $r \geq s$. But E(w) increases by the amount $\eta(s_1 - s)$. Now let s_1 go to infinity. The perturbed distribution function remains within an η-bound of the original one but the mean of the perturbed random variable converges to infinity.

5. Although information restricting $P(y|x,z=1)$ and $P(z|x)$ is superfluous from the perspective of identification, such information may still be useful in practice as it may enable one to obtain more precise sample estimates of $P(y|x,z=1)$ and $P(z|x)$.

6. This classical definition of the treatment effect appropriately characterizes randomized experiments and mandated policies, but other definitions may well be more relevant in many social science applications. For example, one might want to compare exogenous assignment to treatment with self-selection of treatment. A variety of treatment effects of potential interest are considered in Maddala (1983, Section 9.2) and in Heckman and Robb (1985). Our

discussion of identification can easily be extended from the classical treatment effects to these variants.

7. Although it is valid to distinguish mainstream economic thinking on social effects from the perspectives of the other social sciences, one should not think that economists are concerned only with the operation of markets. The field of public economics has long been concerned with "external effects;" social effects on opportunities that operate outside markets. Moreover, some economists have sought to interpret and make use of key sociological ideas. Duesenberry (1949) is one example. More recently, Schelling (1972) analyzed the residential patterns that emerge when individuals choose not to live in neighborhoods where the percentage of residents of their own race is below some threshold. Conlisk (1980) showed that, if decision making is costly, it may be optimal for individuals to imitate the behavior of other persons who are better informed. Akerlof (1980) and Jones (1984) studied the equilibria of noncooperative games in which individuals are punished for deviation from group norms. Gaertner (1974), Pollak (1976), Alessie and Kapteyn (1991), and Case (1991) analyzed consumer demand models in which, holding price fixed, individual demand increases with the mean demand of a reference group.

8. The term "endogenous effects" was introduced in Manski (1991b) to describe a broad class of ideas recurring throughout the social sciences. Sociologists, social psychologists, and some economists

have long been concerned with reinforcing endogenous effects, wherein the propensity of an individual to behave in some way increases with the prevalence of that behavior in the reference group. A host of terms are commonly used to describe reinforcing endogenous effects: "conformity," "imitation," "contagion," "bandwagons," "herd behavior," "norm effects," "keeping up with the Joneses," and, in economics, "interdependent preferences." In addition, economists have always been fundamentally concerned with a particular non-reinforcing endogenous effect: an individual's demand for a product varies with price, which is partly determined by aggregate demand in the relevant market.

9. Inference on contextual effects became an important concern of sociologists in the 1960s, when substantial efforts were made to learn the effects on youth of school and neighborhood environment (e.g. Coleman et al., 1966; Sewell and Armor, 1966). The recent resurgence of interest in spatial concepts of the underclass has spawned many new empirical studies (e.g. Crane, 1991, Jencks and Mayer, 1989, and Mayer, 1991). In Manski (1991b), I use the term "exogenous" effect as a synonym for contextual effect, to distinguish the idea from endogenous effects.

10. The three regressors $E(y|x)$, $E(z|x)$, and x are functionally dependent in the population and so do not vary separately empirically. Nevertheless, one can contemplate the logical

experiment in which one of these regressors is changed and the others held fixed. See the discussion at the end of Section 2.2.3.

11. Beginning with Hyman (1942), reference-group theory has sought to express the idea that individuals learn from or are otherwise influenced by the behavior and attitudes of some reference group. Bank et al. (1990) give an historical account. Sociological writing has remained predominately verbal, but economists have interpreted reference groups as conditioning variables, in the manner of (4.2). See Alessie and Kapteyn (1991) or Manski (1992b).

12. Assumptions (i) and (ii) cover many but not all cases of empirical interest. They are not appropriate in studies of small-group social interactions, such as family interactions. In analyses of family interactions, each reference group (i.e. family) has negligible size relative to the population and random sampling of individuals only rarely yields multiple members of the same family. Hence, it is not a good empirical approximation to assume that x has finite support. Moreover, unless one can characterize groups of families as being similar in composition, it is not plausible to assume that $E(y|x)$ and $E(z|x)$ are continuous functions of x. The conclusion to be drawn, not surprisingly, is that random sampling of individuals is not an effective data gathering process for the study of family interactions. It is preferable to use families as the sampling unit.

13. The same practice is found in empirical studies of decision making under uncertainty. Researchers assume they know the information on which individuals condition their expectations but offer no evidence justifying their assumptions. I have recently criticized this practice in the context of studies of schooling choice. See Manski (1992c).

14. Economists typically assert that respondents to surveys have no incentive to answer questions carefully or honestly; hence, they conclude, there is no reason to think that subjective responses reliably reflect respondents' thinking. Economists' views on the use of subjective data have not, however, always been so negative. In the 1940s, it was common to interview businessmen about their expectations and decision rules. In an influential article, Machlup (1946) sharply attacked existing survey practices as not yielding credible information. This article apparently played an important role in eventually damping the enthusiasm of economists for subjective data. It is revealing that a recent National Academy of Sciences Panel on Survey Measurement of Subjective Phenomena had no economist as a member of the panel and cited almost no economics literature in its report. See Turner and Martin (1984).

15. The impossibility of jointly inferring expectations and preferences from data on oppportunities and choices can easily be seen with a few symbols. The standard economic model assumes that an individual's choice c among specified alternatives C is a

function f(.) of the expected outcomes $(r_i, i \epsilon C)$ associated with the various options; that is, $c = f(r_i, i \epsilon C)$. Suppose that one wishes to learn the decision rule f(.) embodying preferences and mapping expectations into choices. If one observes $\{c, C, (r_i, i \epsilon C)\}$ for a sample of individuals, then one may be able to infer the decision rule. But if one observes only (c,C), then clearly one cannot infer f(.). The most that one can do is infer the decision rule conditional on maintained assumptions on expectations.

References

Ahn, H. and C. Manski(1992), "Distribution Theory for the Analysis of Binary Choice Under Uncertainty with Nonparametric Estimation of Expectations," Journal of Econometrics, forthcoming.

Akerlof, G.(1980), "A Theory of Social Custom, of Which Unemployment may be One Consequence," Quarterly Journal of Economics, 94, 749-775.

Ahn, H. and J. Powell(1992), "Semiparametric Estimation of Censored Selection Models with a Nonparametric Selection Mechanism," Journal of Econometrics, forthcoming.

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

Arabmazar, A. and P. Schmidt(1982), "An Investigation of the Robustness of the Tobit Estimator to Non-normality," Econometrica 50, 1055-1063.

Bank, B., R. Slavings, and B. Biddle(1990), "Effects of Peer, Faculty, and Parental Influences on Students' Persistence," Sociology of Education, 63, 208-225.

Barnhart, J.(1992), letter to Eleanor Chelimsky, in United States General Accounting Office, Unemployed Parents, GAO/PEMD-92-19BR, Gaithersburg, Maryland: U.S. General Accounting Office.

Bassi, L., and O. Ashenfelter (1986), "The Effect of Direct Job Creation and Training Programs on Low-Skilled Workers," in S. Danziger and D. Weinberg (editors), Fighting Poverty, Cambridge, Mass.: Harvard University Press.

Blumstein, A., J. Cohen, and D. Nagin (editors)(1978), Deterrence and Incapacitation: Estimating the Effects of Criminal Sanctions on Crime Rates, Washington, D.C.: National Academy Press.

Borjas, G.(1991), "Ethnic Capital and Intergenerational Mobility," Department of Economics, University of California at San Diego.

Case, A.(1991), "Spatial Patterns in Household Demand," Econometrica, 59, 953-965.

Case, A. and L. Katz (1991), "The Company You Keep: The Effects of Family and Neighborhood on Disadvantaged Youth," Working Paper No. 3705, National Bureau of Economic Research, Cambridge, Mass.

Cliff, A. and J. Ord(1981), Spatial Processes, London: Pion.

Clogg, C.(1992), "The Impact of Sociological Methodology on Statistical Methodology," Statistical Science, 7.

Coleman, J., E. Campbell, C. Hobson, J. McPartland, A. Mood, F. Weinfeld, and R. York(1966), Equality of Educational Opportunity, Washington D.C.: U.S. Government Printing Office.

Conlisk, J.(1980), "Costly Optimizers Versus Cheap Imitators," Journal of Economic Behavior and Organization, 1, 275-293.

Cosslett, S.(1991), "Semiparametric Estimation of a Regression Model with Sample Selectivity," in W. Barnett, J. Powell, and G. Tauchen (editors), Nonparametric and Semiparametric Methods in Econometrics and Statistics, New York: Cambridge University Press.

Coyle, S., R. Boruch, and C. Turner, (editors)(1989), Evaluating AIDS Prevention Programs, Washington, D.C.: National Academy Press.

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

Davidson, A. and L. Beach(1981), "Error Patterns in the Prediction of Fertility Behavior," Journal of Applied Social Psychology, 11, 475-488.

Duesenberry, J.(1949), Income, Savings, and the Theory of Consumption, Cambridge: Harvard University Press.

Duncan, O., A. Haller, and A. Portes(1968), "Peer Influences on Aspirations: A Reinterpretation," American Journal of Sociology, 74, 119-137.

Fishbein, M. and I. Ajzen(1975), Belief, Attitude, Intention, and Behavior: An Introduction to Theory and Research, Reading: Addison-Wesley.

Fisher, F.(1966), The Identification Problem in Econometrics, New York: McGraw-Hill.

Friedman, M.(1957), A Theory of the Consumption Function, Princeton: Princeton University Press.

Frisch, R.(1934), Statistical Confluence Analysis by Means of Complete Regression Systems, Oslo, Norway: University Institute of Economics.

Gaertner, W.(1974), "A Dynamic Model of Interdependent Consumer Behavior," Zeitschrift fur Nationalokonomie, 70, 312-326.

Gamoran, A.(1992),"Social Factors in Education," in M. Alkin (editor), Encyclopedia of Educational Research, 6th Edition, New York: Macmillan, forthcoming.

Gamoran, A. and R. Mare(1989), "Secondary School Tracking and Educational Inequality: Compensation, Reinforcement, or Neutrality?" American Journal of Sociology, 94, 1146-1183.

Goldberger, A.(1983), "Abnormal Selection Bias," in S. Karlin, T. Amemiya, and L. Goodman (editors), Studies in Econometrics, Time Series, and Multivariate Statistics, Orlando: Academic Press.

Gronau, R.(1974), "Wage Comparisons - a Selectivity Bias," Journal of Political Economy, 82, 1119-1143.

Hanushek, E.(1986), "The Economics of Schooling," Journal of Economic Literature, 24, 1141-1177.

Hardle, W.(1990), Applied Nonparametric Regression, Cambridge, England: Cambridge University Press.

Hauser, R.(1970), "Context and Consex: A Cautionary Tale," American Journal of Sociology, 75, 645-664.

Heckman, J.(1976), "The Common Structure of Statistical Models of Truncation, Sample Selection, and Limited Dependent Variables and a Simple Estimator for Such Models," Annals of Economic and Social Measurement, 5, 479-492.

Heckman, J.(1978), "Dummy Endogenous Variables in a Simultaneous Equation System," Econometrica 46, 931-959.

Heckman, J. and B. Honore,(1990), "The Empirical Content of the Roy Model," Econometrica 58, 1121-1149.

Heckman, J. and R. Robb(1985), "Alternative Methods for Evaluating the Impact of Interventions," in J. Heckman and B. Singer (editors), Longitudinal Analysis of Labor Market Data, Cambridge: Cambridge University Press.

Hendershot, G. and P. Placek (editors)(1981), Predicting Fertility, Lexingon: D. C. Heath.

Hayes, C. and S. Hofferth (editors)(1987), Risking the Future: Adolescent Sexuality, Pregnancy, and Childbearing, Washington, D.C.: National Academy Press.

Huber, P.(1981), Robust Statistics, New York: Wiley.

Hurd, M.(1979), "Estimation in Truncated Samples When There is Heteroskedasticity," Journal of Econometrics 11, 247-258.

Hyman, H.(1942), "The Psychology of Status," Archives of Psychology," No. 269.

Ichimura, Hidehiko, and Lee, Lung-Fei(1991), "Semiparametric Estimation of Multiple Indices Models: Single Equation Estimation," in W. Barnett, J. Powell, and G. Tauchen (editors), Nonparametric and Semiparametric Methods in Econometrics and Statistics, Cambridge, England: Cambridge University Press.

Jencks, C. and S. Mayer(1989), "Growing Up in Poor Neighborhoods: How Much Does it Matter?" Science, 243, 1441-1445.

Jones, S.(1984), The Economics of Conformism, Oxford: Basil Blackwell.

Juster, T.(1966), "Consumer Buying Intentions and Purchase Probability: An Experiment in Survey Design," Journal of the American Statistical Association, 61, 658-696.

Kalbfleisch, J. and Prentice, R.(1980), The Statistical Analysis of Failure Time Data, New York: Wiley.

Koenker, R. and Bassett, G.(1978), "Regression Quantiles," Econometrica, 46, 33-50.

LaLonde, R.(1986), "Evaluating the Econometric Evaluations of Training Programs with Experimental Data," American Economic Review, 76, 604-620.

Machlup, F. (1946), "Marginal Analysis and Empirical Research," American Economic Review, 36, 519-554.

Maddala, G.S.(1983), Qualitative and Limited Dependent Variable Models in Econometrics, Cambridge, England: Cambridge University Press.

Manski, C.(1981), "Structural Models for Discrete Data: The Analysis of Discrete Choice," in S. Leinhardt (editor), Sociological Methodology 1981, San Francisco: Jossey-Bass.

Manski, C.(1988a), Analog Estimation Methods in Econometrics, London, England: Chapman and Hall.

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

Manski, C.(1989), "Anatomy of the Selection Problem," Journal of Human Resources, 24, 343-360.

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

Manski, C.(1990b), "The Use of Intentions Data to Predict Behavior: A Best Case Analysis," Journal of the American Statistical Association, 85, 934-940.

Manski, C.(1991a), "Regression,"Journal of Economic Literature, 29, 34-50.

Manski, C.(1991b), "Identification of Endogenous Social Effects: the Reflection Problem," Social Systems Research Institute Discussion Paper 9127, University of Wisconsin-Madison.

Manski, C. (1992a), "The Selection Problem," in C. Sims (editor), Advances in Econometrics: Sixth World Congress of the Econometric Society, New York: Cambridge University Press.

Manski, C.(1992b), "Dynamic Choice in a Social Setting: Learning from the Experiences of Others," Journal of Econometrics, forthcoming.

Manski, C.(1992c), "Adolescent Econometricians: How Do Youth Infer the Returns to Schooling?" in C. Clotfelter and M. Rothschild (editors), The Economics of Higher Education, Chicago: University of Chicago Press, forthcoming.

Manski, C. and Garfinkel, I. (editors)(1992), Evaluating Welfare and Training Programs, Cambridge, Mass.: Harvard University Press.

Manski, C. and McFadden, D.(1981), "Alternative Estimators and Sample Designs for Discrete Choice Analysis," in C. Manski and D. McFadden (editors), Structural Analysis of Discrete Data with Econometric Applications, Cambridge, Mass.: M.I.T. Press.

Manski, C., G. Sandefur, S. McLanahan, and D. Powers (1992), "Alternative Estimates of the Effect of Family Structure During Adolescence on High School Graduation," Journal of the American Statistical Association, 87, 25-37.

Manski, C. and D. Wise(1983), College Choice in America, Cambridge: Harvard University Press.

Mayer, S.(1991), "How Much Does a High School's Racial and Socioeconomic Mix Affect Graduation and Teenage Fertility Rates?" in C. Jencks and P. Peterson (editors) The Urban Underclass, Washington, D.C.: The Brookings Institution.

Moffitt, R.(1992), "Incentive Effects of the U.S. Welfare System: A Review," Journal of Economic Literature, 30, 1-61.

O'Connell, M. and C. Rogers(1983), "Assessing Cohort Birth Expectations Data from the Current Population Survey, 1971-1981," Demography 20, 369-383.

Pollak, R.(1976), "Interdependent Preferences," American Economic Review, 78, 745-763.

Robinson, C.(1989), "The Joint Determination of Union Status and Union Wage Effects: Some Tests of Alternative Models," Journal of Political Economy, 97, 639-667.

Robinson, P.(1988), "Root-N-Consistent Semiparametric Regression," Econometrica, 56, 931-954.

Rubin, D.(1987), Multiple Imputation for Nonresponse in Surveys, New York: Wiley.

Schelling, T. (1971), "Dynamic Models of Segregation," Journal of Mathematical Sociology, 1, 143-186.

Sewell, W. and J. Armer(1966), "Neighborhood Context and College Plans," American Sociological Review, 31, 159-168.

Stigler, S.(1986), The History of Statistics, Cambridge, Mass: Belknap Press.

Tobin, J.(1958), "Estimation of Relationships for Limited Dependent Variables," Econometrica, 26, 24-36.

Turner, C. and E. Martin (editors)(1984), Surveying Subjective Phenomena, volume 1, New York: Russell Sage Foundation.

U.S. Bureau of the Census(1988), Current Population Reports, Series P-20, No. 427, Fertility of American Women: June, 1987, Washington, D.C.: U.S. Government Printing Office.

Wainer, H. (editor)(1986), Drawing Inferences from Self-Selected Samples, New York: Springer.

Wainer, H. (editor)(1989), Journal of Educational Statistics, 14(2).

Westoff, C. and N. Ryder(1977), "The Predictive Validity of Reproductive Intentions," Demography 14, 431-453.

Winship, C. and R. Mare(1992), "Models for Sample Selection Bias," Annual Review of Sociology, 18.

Woittiez, I. and A. Kapteyn(1991), "Social Interactions and Habit Formation in a Labor Supply Model," Department of Economics, University of Leiden.

RECENT SSRI WORKING PAPERS

9127
Manski, Charles F.
IDENTIFICATION OF ENDOGENOUS SOCIAL EFFECTS: THE REFLECTION PROBLEM

9201
LeBaron, Blake
PERSISTENCE OF THE DOW JONES INDEX ON RISING VOLUME

9202
Berkowitz, Daniel and Beth Mitchneck
FISCAL DECENTRALIZATION IN THE SOVIET ECONOMY

9203
Streufert, Peter A.
A GENERAL THEORY OF SEPARABILITY FOR PREFERENCES DEFINED ON A COUNTABLY INFINITE PRODUCT SPACE

9204
Baek, Ehung G. and William A. Brock
A NONPARAMETRIC TEST FOR INDEPENDENCE OF A MULTIVARIATE TIME SERIES

9205
Mailath, George J., Larry Samuelson and Jeroen M. Swinkels
NORMAL FORM STRUCTURES IN EXTENSIVE FORM GAMES

9206
Mirman, Leonard J., Larry Samuelson and Amparo Urbano
DUOPOLY SIGNAL JAMMING

9207
Andreoni, James and Ted Bergstrom
DO GOVERNMENT SUBSIDIES INCREASE THE PRIVATE SUPPLY OF PUBLIC GOODS?

9208
Loretan, Mico and Peter C.B. Phillips
TESTING THE COVARIANCE STATIONARITY OF HEAVY-TAILED TIME SERIES: AN OVERVIEW OF THE THEORY WITH APPLICATIONS TO SEVERAL FINANCIAL DATASETS

9209
Horowitz, Joel L. and Charles F. Manski
IDENTIFICATION AND ROBUSTNESS IN THE PRESENCE OF ERRORS IN DATA

9210
Mirman, Leonard J., Larry Samuelson and Edward E. Schlee
STRATEGIC INFORMATION MANIPULATION IN DUOPOLIES

9211
Holmes, Thomas J. and James A. Schmitz, Jr.
ON THE TURNOVER OF BUSINESS FIRMS AND BUSINESS MANAGERS

9212
Che, Yeon-Koo
THE ECONOMICS OF COLLECTIVE NEGOTIATION: CONSOLIDATION OF CLAIMS BY MULTIPLE PLAINTIFFS

9213
Phelan, Christopher
INCENTIVES, INEQUALITY, AND THE BUSINESS CYCLE

9214
Gale, Ian
PRICE DISPERSION IN A MARKET WITH ADVANCE PURCHASES

9215
Gale, Ian and Donald Hausch
BOTTOM-FISHING AND DECLINING PRICES IN SEQUENTIAL AUCTIONS

9216
Barham, Bradford, Michael Carter and Wayne Sigelko
ADOPTION AND ACCUMULATION PATTERNS IN GUATEMALA'S LATEST AGRO-EXPORT BOOM

9217
Manski, Charles F.
IDENTIFICATION PROBLEMS IN THE SOCIAL SCIENCES

This Series includes selected publications of members of SSRI and others working in close association with the Institute. Inquiries concerning publications should be addressed to Ann Teeters Lewis, SSRI, 6470 Social Science Bldg., University of Wisconsin-Madison, Madison, WI 53706. (608)-263-3876. Electronic mail address SSRI@vms.macc.wisc.edu.