TECHNICAL WORKING PAPER SERIES

THE MIXING PROBLEM IN PROGRAM EVALUATION

Charles F. Manski

Technical Working Paper No. 148

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
December 1993

This research was undertaken while the author was a Fellow at the Center for Advanced Study in the Behavioral Sciences, supported by National Science Foundation Grant SES90-22192 and by a grant from the University of Wisconsin Graduate School. The work is also supported by National Science Foundation Grant SES92-23220. Earlier drafts (dated March, April, and August 1993), were circulated under the title "What Do Controlled Experiments Reveal About Outcomes When Treatments Vary?". I have benefitted from extended discussions with Joe Hotz, Robert Moffitt, and Yu Xie, as well as from comments received in seminar and conference presentations at the University of California at Santa Barbara, RAND, the Hebrew University of Jerusalem, the University of Udine, NBER, and CEMFI. This paper is part of NBER's research program in Labor Studies. Any opinions expressed are those of the author and not those of the National Bureau of Economic Research.

NBER Technical Working Paper #148
December 1993

THE MIXING PROBLEM IN PROGRAM EVALUATION

ABSTRACT

A common concern of evaluation studies is to learn the distribution of outcomes when a specified *treatment policy*, or assignment rule, determines the treatment received by each member of a specified population. Recent studies have emphasized evaluation of policies providing the same treatment to all members of the population. In particular, experiments with randomized treatments have this objective. Policies mandating homogenous treatment of the population are of interest, but so are ones that permit treatment to vary across the population. This paper examines the use of empirical evidence on programs with homogenous treatments to infer the outcomes that would occur if treatment were to vary across the population. Experimental evidence from the Perry Preschool Project is used to illustrate the inferential problem and the main findings of the analysis.

Charles F. Manski
Department of Economics
University of Wisconsin-Madison
1180 Observatory Drive
Madison, WI 53706
and NBER

1. Introduction

A common concern of evaluation studies is to learn the distribution of outcomes when a specified *treatment policy*, or assignment rule, determines the treatment received by each member of a specified population. Recent studies have emphasized evaluation of policies providing the same treatment to all members of the population. In particular, experiments with randomized treatments have this objective. The classical experimental protocol calls for random samples of the population to be drawn and formed into treatment groups, all of whose members are assigned the same treatment. The empirical distribution of outcomes realized by a treatment group is then ostensibly the same (up to random sampling error) as would be observed if the treatment in question were applied to the entire population. For example, see Manski and Garfinkel (1992), some of whose chapters describe recent experimental evaluations of welfare and training programs.

Policies mandating homogeneous treatment of the population are of interest, but so are ones that permit treatment to vary across the population. We often see policies calling on persons to select their own treatments. Policies intended to mandate homogeneous treatment sometimes turn out to be voluntary in practice, as compliance with the mandated treatment is not enforced. Resource constraints sometimes prevent universal implementation of desirable treatments.

Consider the following inferential questions:

* What do observations of outcomes when treatments vary across the population reveal about the outcomes that would occur if treatment were homogeneous?

** What do observations of outcomes when treatment is homogeneous reveal about the outcomes that would occur if treatment were to vary across the population?

The first question, usually called the *selection* or *switching* problem, has drawn considerable attention and much has been learned; see Maddala (1983), Heckman and Robb (1985), and Manski (1993). The second question, which has remained unexplored and unnamed, is the subject of this paper. Formally, the question asks what inferences about mixtures of two random variables can be made given knowledge of their marginal distributions. Hence I refer to it as the *mixing* problem.

SELECTION AND MIXING PROBLEMS: To formalize these inferential questions, let each member of the population be described by values for $[(y_1,y_0),z_m,x]$. Here x is a vector of covariates, an element of some space X. There are two feasible treatments, labelled 1 and 0.[1] The letter m denotes the treatment policy of interest. A treatment policy determines which treatment each person receives. The indicator variable z_m denotes the treatment that a given person receives under policy m; $z_m = 1$ if the person receives treatment 1 and $z_m = 0$ otherwise. Associated with the treatments are outcomes (y_1,y_0), a pair of elements of some outcome space Y. The outcome a person realizes under policy m is

(1) $y_m \equiv y_1 z_m + y_0(1-z_m)$.

The distribution of outcomes realized by those persons sharing the same value of x is

$$(2)\quad P(y_m \mid x) \equiv P[y_1 z_m + y_0(1-z_m) \mid x]$$
$$= P(y_1 \mid x, z_m=1)P(z_m=1 \mid x) + P(y_0 \mid x, z_m=0)P(z_m=0 \mid x).$$

For example, a welfare recipient might be treated by job-specific training or by basic education. The relevant outcome might be earned income following treatment. One treatment policy might mandate job training for all welfare recipients and enforce the mandate. A second policy might attempt to mandate basic education but not be able to enforce compliance. A third policy might permit a person's case worker to select the treatment expected to yield the larger net benefit, measured as earned income minus treatment costs.

The problem of interest is to learn about the distribution $P(y_m \mid x)$ of outcomes that would be realized by persons with covariates x if a specified treatment policy m were in effect. Inference is straightforward if one can enact policy m and observe the outcomes. The interesting inferential questions concern the feasibility of learning $P(y_m \mid x)$ when one observes outcomes under policies other than m. The selection problem and the mixing problem both concern the the feasibility of extrapolating from observed treatment policies to unobserved ones.

The selection problem arises when policy m mandates homogeneous treatment, but the available data are realizations under some other policy that may yield heterogeneous treatments. Suppose that m makes treatment 1 mandatory for all persons with covariates x, so $P(z_m=1 \mid x) = 1$ and $P(y_m \mid x) = P(y_1 \mid x)$. Suppose that the observable policy is some $\mu \neq m$.[2] The sampling process identifies the censored outcome distributions $P(y_1 \mid x, z_\mu=1)$ and

$P(y_0 \mid x, z_\mu = 0)$, as well as the treatment distribution $P(z_\mu \mid x)$. So the formal statement of the selection problem is:

* What does knowledge of $[P(y_1 \mid x, z_\mu = 1), P(y_0 \mid x, z_\mu = 0), P(z_\mu \mid x)]$ imply about $P(y_1 \mid x)$?

The mixing problem arises when policy m may yield heterogenous treatments, but the available data are realizations under policies imposing homogenous treatments. In particular, the classical model of experimentation presumes that experimental evidence is available for both treatments, so the experiments identify $P(y_1 \mid x)$ and $P(y_0 \mid x)$. So the formal statement of the mixing problem is:[3]

** What does knowledge of $[P(y_1 \mid x), P(y_0 \mid x)]$ imply about $P[y_1 z_m + y_0(1-z_m) \mid x]$?

ORGANIZATION OF THE PAPER: Section 2 uses empirical evidence from a famous social experiment, the Perry Preschool Project, to illustrate the mixing problem and the main findings of this paper. Fifteen years after their participation in this early-childhood educational intervention, sixty-seven percent of a treatment group were high school graduates. At the same time, only forty-nine percent of a control group were graduates. Our interest is to determine what the experimental evidence and various forms of prior information imply about the rate of high school graduation that would prevail under treatment policies applying the intervention to some children but not to others.

Sections 3 through 6 present the analysis yielding the empirical results reported in Section 2. In my earlier study of the selection problem (Manski, 1989, 1993), I found it productive to begin by determining what can be learned when the sampling process provides the only information available to the researcher. I then examined the identifying power of various forms of prior information that might plausibly be invoked in empirical studies. The present analysis uses the same approach.

Section 3 investigates the mixing problem when knowledge of the two marginal distributions $P(y_1 \mid x)$ and $P(y_0 \mid x)$ is the only information available. The basic finding is a proposition giving sharp bounds on conditional probabilities of the form $P(y_m \epsilon B \mid x)$, $B \subset Y$. When outcomes are real-valued, this finding is easily transformed into sharp bounds on the quantiles of $P(y_m \mid x)$.[4]

The bounds of Section 3 may be tightened if the researcher possesses prior information on the distribution of $[(y_1,y_0),z_m,x]$. Section 4 examines the identifying power of information restricting the joint distribution of outcomes $P(y_1,y_0 \mid x)$.[5] Section 4.1 assumes that y_1 and y_0 are statistically independent, conditional on the covariates x. In contrast, Section 4.2 supposes that the outcomes are shifted versions of one another, with $y_1 = y_0 + \delta$ for some constant δ. Section 4.3 assumes that outcomes are ordered, with $y_1 \geq y_0$ for all persons with covariates x.

Section 5 examines restrictions on the distribution $P[z_m \mid (y_1,y_0),x]$ describing treatment policy m. Section 5.1 assumes that the treatment z_m received by each person is statistically independent of that person's outcomes (y_1,y_0), conditional on x. Section 5.2 supposes that the treatment policy minimizes or maximizes the probability that y_m falls in specified sets of events; important applications include the Roy model and the competing risks model. Section 5.3

assumes that one knows the size $P(z_m=1 \mid x)$ of the subpopulation receiving treatment 1, but does not know the composition of this subpopulation. An interesting finding is that knowledge of $P(z_m=1 \mid x)$ makes it possible to learn something about $P(y_m \mid x)$ even if one of the two distributions $[P(y_1 \mid x), P(y_0 \mid x)]$ is not known.

Taken one at a time, each of the assumptions imposed in Section 4 and 5 on the distribution of outcomes or on the treatment policy implies a distinctive bound on $P(y_m \mid x)$, but none of the assumptions is strong enough to identify the distribution. Combinations of assumptions do identify $P(y_m \mid x)$. Two such are stated in Section 6.

IDENTIFICATION AND SAMPLE INFERENCE: The mixing problem, like the selection problem, is a failure of identification rather than a difficulty in sample inference. To keep attention focussed on identification, Sections 3 through 6 maintain the assumption that the marginal distributions $P(y_1 \mid x)$ and $P(y_0 \mid x)$ are known almost everywhere on the covariate space. The identification findings reported in these sections can be translated into consistent sample estimates of identified quantities by replacing $P(y_1 \mid x)$ and $P(y_0 \mid x)$ with consistent nonparametric estimates, as is done in Section 2. Moreover, sampling confidence bands can be placed around these estimates, much as they were by Manski et al. (1992) in an empirical study concerned with the selection problem.

For the sake of simplicity, I often refer to $P(y_1 \mid x)$ and $P(y_0 \mid x)$ simply as the distributions of y_1 and y_0, rather than as distributions conditional on x. One could similarly shorten the notation by denoting these distributions as $P(y_1)$ and $P(y_0)$. I do not take this step because I want the reader to keep in mind that the analysis of this paper holds for any specification of the covariates x.

CAVEATS ON CLASSICAL EXPERIMENTATION: Where it discusses experimentation with randomized treatments, this paper maintains the classical assumption that experimental regimes operate exactly as would mandatory treatment policies. I have elsewhere discussed some of the many reasons why this central tenet of experimental analysis may fail to hold when applied to welfare and training programs (see the introduction to Manski and Garfinkel, 1992). Experiments may be administered differently from actual programs. Macro feedback effects ranging from information diffusion to norm formation to market equilibration may make the full-scale implementation of a treatment policy inherently different from the small-scale implementation of an experiment. Strictures on forcing human subjects into experiments may make it impossible to form random treatment groups. It may not be practical to execute experiments covering more than a small subset of the treatments and environments that are germane to policy formation. The present analysis assumes away all of these very real concerns in order to focus on the mixing problem.

2. An Illustration: The Perry Preschool Project

Beginning in 1962, the Perry Preschool Project provided intensive educational and social services to a random sample of black children in Ypsilanti, Michigan. The project investigators also drew a second random sample of such children, but provided them with no special services. Subsequently, a variety of outcomes were ascertained for most members of the treatment and control groups. Among other things, it was found that sixty-seven percent of the treatment group and forty-nine percent of the control group were high school graduates by age 19 (see Berrueta-Clement et al., 1984). This and similar findings for other outcomes have been widely cited as evidence that intensive early childhood educational interventions improve the outcomes of children at risk (see Holden, 1990).

For purposes of discussion, let us accept the Perry Preschool Project as a classical experiment, with

x = black children in Ypsilanti, Michigan

$y_1 = 1$ if high school graduate by age 19, $= 0$ otherwise; intervention received.

$y_0 = 1$ if high school graduate by age 19, $= 0$ otherwise; intervention not received.

Moreover, ignoring attrition and sampling error in the estimation of outcome distributions, let us accept the experimental evidence as showing that the high school graduation rate among children with covariate value x would be .67 if all such children were to receive the intervention, and would be .49 if none of them were to receive the intervention. That is, let us accept the experimental evidence as showing that $P(y_1=1 \mid x) = .67$ and $P(y_0=1 \mid x) = .49$.[6]

What would be the rate $P(y_m=1 \mid x)$ of high school graduation under a treatment policy m where some children with covariates x receive the intervention, but not others? Table 1 summarizes the inferences that can be made given the experimental evidence and varying forms of prior information about the outcome distribution and the treatment policy. In each case, the table cites a proposition implying the estimate shown. These propositions are developed in Sections 3 through 6.

IDENTIFICATION USING ONLY THE EXPERIMENTAL EVIDENCE: It might be conjectured that $P(y_m=1 \mid x)$ must lie between the graduation rates of the control and treatment groups, namely .49 and .67. This conjecture is correct for special outcome distributions and treatment policies. It holds if

(a) the outcomes (y_1,y_0) are ordered, with $y_1 \geq y_0$ for all children

or if

(b) the treatment policy makes z_m statistically independent of the outcomes (y_1,y_0).

The conjecture does not hold more generally. In fact, the experimental evidence only implies that the graduation rate must lie between .16 and 1. That is, there exist outcome distributions and treatment policies that are consistent with the known values of $P(y_1 \mid x)$ and $P(y_0 \mid x)$ and that imply graduation rates as low as .16 and as high as 1.

This result is easily understood once one considers precisely what the experimental evidence does and does not reveal. Observing the outcomes of the treatment group reveals (ignoring sampling error) that $y_1 = 1$ for 67 percent of the population and $y_1 = 0$ for the remaining 33 percent. Observing the outcomes of the control group reveals that $y_0 = 1$ for 49 percent of the population and $y_0 = 0$ for the remaining 51 percent.

The experimental evidence does not reveal how y_1 and y_0 are related within the population, nor how policy m assigns treatments. The impact of treatment policy on the graduation rate is most pronounced when y_1 and y_0 are most negatively associated. Among all distributions of (y_1,y_0) that are consistent with the experimental evidence, the one with the greatest negative association between y_1 and y_0 is this:

$P(y_1=0,y_0=0 \mid x) = .00 \quad P(y_1=0,y_0=1 \mid x) = .33$

$P(y_1=1,y_0=0 \mid x) = .51 \quad P(y_1=1,y_0=1 \mid x) = .16.$

Given this distribution of outcomes, the graduation rate is maximized by adopting a treatment policy that gives the intervention only to those children with $y_1 = 1$. The result is a 100 percent graduation rate. At the other extreme, the graduation probability is minimized by adopting a treatment policy that gives the intervention only to those children with $y_1 = 0$. The result is a 16 percent graduation rate.

PRIOR INFORMATION: The interval [.16,1] is a "worst-case" bound on the graduation rate, computed in the absence of any prior information restricting the outcome distribution or the treatment policy. A researcher who possesses such information may be able to narrow the range of possible graduation rates.

Imagine that one has no information about the treatment policy but does have information about the outcome distribution. One might think that being treated by the preschool intervention can never harm a child's schooling prospects; that is, outcomes are ordered with $y_1 \geq y_0$ for all children. If so, then the graduation rate must lie between those observed in the control and treatment groups, namely .49 and .67. A more neutral assumption might be that y_1 and y_0 are statistically independent conditional on x. This assumption implies that the graduation rate must lie between .33 and .83; where the rate falls within this range depends on the treatment policy.

Next imagine that one has no information about the outcome distribution but does have information about the treatment policy. One might think that treatment decisions will be made by omniscient parents who choose for each child the treatment yielding the better outcome. This assumption implies that the graduation rate must lie between .67 and 1; where the rate falls within this range depends on the outcome distribution. On the other hand, one might think that assignments to treatments are statistically independent of outcomes, as they would be if an explicit random assignment rule is used. Then the graduation rate must lie between the .49 and .67 observed in the control and treatment groups.

Finally, imagine that resource constraints limit implementation of the intervention to part of the population. Suppose that one knows the fraction of the population receiving the intervention, but does not know the composition of the treated and untreated subpopulations. As Table 1 shows, knowing that 1/10 or 5/10 or 9/10 of the population receives the intervention implies that the graduation rate must lie in the interval [.39,.59] or [.17,.99] or [.57,77] respectively. Observe that the first and third intervals are relatively narrow but the second is rather wide, almost as wide as the interval found in the absence of prior information. This pattern of results reflects the fact that the power of treatment policy to determine who receives which treatment is much more constrained when $P(z_m=1 \mid x)$ is fixed at a value near zero or one than it is when $P(z_m=1 \mid x)$ is fixed at 5/10.

The scenarios considered thus far bring to bear enough empirical evidence and prior information to bound the high school graduation rate but not to identify it. If stronger restrictions are imposed, then the high school graduation rate may be identified. For example, if it is known that the outcomes (y_1, y_0) are statistically independent and that each child receives the treatment yielding the better outcome, then the implied high school graduation rate is .83. If it is known that 5/10 of the population receives the intervention and that treatment is independent of outcomes, then the implied graduation rate is .58.

The general lesson is that experimental evidence alone permits only weak conclusions to be drawn about the high school graduation rate when treatments vary. Experimental evidence combined with prior information implies stronger conclusions. The nature of these stronger conclusions depends critically on the prior information asserted. This lesson is analogous to the one learned over the past twenty years about the conclusions that can be drawn about mandatory programs from observations of outcomes when treatments vary. Mixing and selection are distinct identification problems, but they are closely related.

TABLE I: THE PERRY PRESCHOOL PROJECT

Experimental Evidence

$P(y_1=1 \mid x) = .67 \quad P(y_0=1 \mid x) = .49$

Prior Information	$P(y_m=1 \mid x)$
no prior information (Proposition 1)	[.16,1]
independent outcomes (Proposition 2)	[.33,.83]
ordered outcomes (Proposition 4)	[.49,.67]
treatment independent of outcomes (Proposition 5)	[.49,.67]
treatment maximizing graduation rate (Proposition 6B)	[.67,1]
+ independent outcomes (Proposition 9B)	.83
+ ordered outcomes (Proposition 9B)	.67
treatment minimizing graduation rate (Proposition 6A)	[.16,.49]
1/10 population receives treatment 1 (Proposition 7)	[.39,.59]
+ treatment independent of outcomes (Proposition 9A)	.51
5/10 population receives treatment 1 (Proposition 7)	[.17,.99]
+ treatment independent of outcomes (Proposition 9A)	.58
9/10 population receives treatment 1 (Proposition 7)	[.57,.77]
+ treatment independent of outcomes (Proposition 9A)	.65

3. Identification of Mixtures Using Only Knowledge of The Marginals

This section characterizes the restrictions on $P(y_m \mid x)$ implied by knowledge of $[P(y_1 \mid x), P(y_0 \mid x)]$. No other information is assumed available.

PROBABILITIES OF EVENTS: Consider the probability that the realized outcome y_m falls in some set B, conditional on x; that is, $P(y_m \epsilon B \mid x)$. Given that y_m always equals either y_1 or y_0, one might think that $P(y_m \epsilon B \mid x)$ must lie between $P(y_1 \epsilon B \mid x)$ and $P(y_0 \epsilon B \mid x)$. This is not the case. It turns out that when $P(y_1 \epsilon B \mid x) + P(y_0 \epsilon B \mid x) \leq 1$, then $P(y_m \epsilon B \mid x)$ must lie in the interval $[0, P(y_1 \epsilon B \mid x) + P(y_0 \epsilon B \mid x)]$. When $P(y_1 \epsilon B \mid x) + P(y_0 \epsilon B \mid x) \geq 1$, $P(y_m \epsilon B \mid x)$ must lie in the interval $[P(y_1 \epsilon B \mid x) + P(y_0 \epsilon B \mid x) - 1, 1]$. Proposition 1 gives the result.

Proposition 1: Let $P(y_1 \mid x)$ and $P(y_0 \mid x)$ be known. Then

$$(3)\quad \max[0, P(y_1 \epsilon B \mid x) + P(y_0 \epsilon B \mid x) - 1] \leq P(y_m \epsilon B \mid x) \leq \min[P(y_1 \epsilon B \mid x) + P(y_0 \epsilon B \mid x), 1]. \quad \blacksquare$$

PROOF: We first determine the treatment policies that minimize and maximize $P(y_m \epsilon B \mid x)$. Observe that if y_1 and y_0 both fall in the set B, then y_m must fall in B. Moreover, if neither y_1 nor y_0 falls in B, then y_m cannot fall in B. That is,

$$(4a)\quad y_1 \epsilon B \cap y_0 \epsilon B \Rightarrow y_m \epsilon B$$

and

(4b) $y_1 \notin B \cap y_0 \notin B \Rightarrow y_m \notin B$,

whatever treatment policy m may be.

The treatment policy is relevant only in those cases where one of the two outcomes falls in B and the other does not. The treatment policy minimizes $P(y_m \in B \mid x)$ if it always selects the treatment yielding the outcome not in B; that is, if

(5) $y_1 \notin B \cap y_0 \in B \Rightarrow z_m = 1$

$y_1 \in B \cap y_0 \notin B \Rightarrow z_m = 0$.

Hence, the smallest possible value of $P(y_m \in B \mid x)$ is $P(y_1 \in B \cap y_0 \in B \mid x)$. The treatment policy maximizes $P(y_m \in B \mid x)$ if it always selects the treatment yielding the outcome in B; that is, if

(6) $y_1 \notin B \cap y_0 \in B \Rightarrow z_m = 0$

$y_1 \in B \cap y_0 \notin B \Rightarrow z_m = 1$.

So the largest possible value of $P(y_m \in B \mid x)$ is $P(y_1 \in B \cup y_0 \in B \mid x)$.

The above shows that if $P(y_1 \in B \cap y_0 \in B \mid x)$ and $P(y_1 \in B \cup y_0 \in B \mid x)$ are known, then

(7) $P(y_1 \in B \cap y_0 \in B \mid x) \le P(y_m \in B \mid x) \le P(y_1 \in B \cup y_0 \in B \mid x)$

is the sharp bound on $P(y_m \epsilon B \mid x)$. But the only available information is knowledge of $P(y_1 \mid x)$ and $P(y_0 \mid x)$. Therefore, the best computable lower bound on $P(y_m \epsilon B \mid x)$ is the smallest value of $P(y_1 \epsilon B \cap y_0 \epsilon B \mid x)$ that is consistent with the known $P(y_1 \mid x)$ and $P(y_0 \mid x)$. Similarly, the best computable upper bound is the largest feasible value of $P(y_1 \epsilon B \cup y_0 \epsilon B \mid x)$.

The second step is to determine these best computable bounds. This is simple to do, because Frechet (1951) proved this sharp bound on $P(y_1 \epsilon B \cap y_0 \epsilon B \mid x)$, given knowledge of $P(y_1 \mid x)$ and $P(y_0 \mid x)$:[7]

$$(8) \quad \max[0, P(y_1 \epsilon B \mid x) + P(y_0 \epsilon B \mid x) - 1] \leq P(y_1 \epsilon B \cap y_0 \epsilon B \mid x) \leq \min[P(y_1 \epsilon B \mid x), P(y_0 \epsilon B \mid x)].$$

It follows immediately from (8) that the best computable lower bound on $P(y_m \mid x)$ is $\max[0, P(y_1 \epsilon B \mid x) + P(y_0 \epsilon B \mid x) - 1]$. To obtain the best computable upper bound, observe that

$$(9) \quad P(y_1 \epsilon B \cup y_0 \epsilon B \mid x) = P(y_1 \epsilon B \mid x) + P(y_0 \epsilon B \mid x) - P(y_1 \epsilon B \cap y_0 \epsilon B \mid x).$$

Applying the Frechet lower bound on $P(y_1 \epsilon B \cap y_0 \epsilon B \mid x)$ to (9) shows that

$$(10) \quad P(y_1 \epsilon B \cup y_0 \epsilon B \mid x) \leq \min[P(y_1 \epsilon B \mid x) + P(y_0 \epsilon B \mid x), 1].$$

Hence $\min[P(y_1 \epsilon B \mid x) + P(y_0 \epsilon B \mid x), 1]$ is the best computable upper bound on $P(y_m \mid x)$.

Q.E.D.

QUANTILES: Suppose that Y is the real line. Let $u \in R^1$ and $B = (-\infty, u]$. By Proposition 1,

$$(11) \quad \max[0, P(y_1 \leq u \mid x) + P(y_0 \leq u \mid x) - 1] \leq P(y_m \leq u \mid x)$$
$$\leq \min[P(y_1 \leq u \mid x) + P(y_0 \leq u \mid x), 1].$$

Let $\alpha \in (0,1)$ and let $q_m(\alpha \mid x)$ denote the α-quantile of y_m, conditional on x. Corollary 1.1 inverts the bound (11) to obtain a sharp bound on $q_m(\alpha \mid x)$.

Corollary 1.1: Let $P(y_1 \mid x)$ and $P(y_0 \mid x)$ be known. Let Y be the real line. Let

$$r_1(\alpha \mid x) \equiv \inf_u \text{ s.t. } P(y_1 \leq u \mid x) + P(y_0 \leq u \mid x) \geq \alpha$$

$$s_1(\alpha \mid x) \equiv \inf_u \text{ s.t. } P(y_1 \leq u \mid x) + P(y_0 \leq u \mid x) - 1 \geq \alpha.$$

Then

$$(12) \quad r_1(\alpha \mid x) \leq q_m(\alpha \mid x) \leq s_1(\alpha \mid x). \quad \blacksquare$$

PROOF: By the upper bound on $P(y_m \leq u \mid x)$ in (11),

$$u < r_1(\alpha \mid x) \Rightarrow P(y_1 \leq u \mid x) + P(y_0 \leq u \mid x) < \alpha$$
$$\Rightarrow P(y_m \leq u \mid x) < \alpha$$
$$\Rightarrow q_m(\alpha \mid x) > u.$$

Hence, $r_1(\alpha \mid x) \leq q_m(\alpha \mid x)$. By the lower bound on $P(y_m \leq u \mid x)$ in (11),

$$u \geq s_1(\alpha \mid x) \Rightarrow P(y_1 \leq u \mid x) + P(y_0 \leq u \mid x) - 1 \geq \alpha$$
$$\Rightarrow P(y_m \leq u \mid x) \geq \alpha$$
$$\Rightarrow q_m(\alpha \mid x) \leq u.$$

Hence, $q_m(\alpha \mid x) \leq s_1(\alpha \mid x)$. These bounds on $q_m(\alpha \mid x)$ are sharp because the bounds in (11) are sharp.

Q.E.D.

It is interesting that these bounds on quantiles of $P(y_m \mid x)$ are always informative both above and below. This is so even though the bound on $P(y_m \leq u \mid x)$ used to derive Corollary 1.1 is only informative above or below, the informative direction depending on the value of u.

4. Restrictions on the Outcome Distribution

In the course of proving Proposition 1, we showed that if $P(y_1 \epsilon B \cap y_0 \epsilon B \mid x)$ and $P(y_1 \epsilon B \cup y_0 \epsilon B \mid x)$ are known and if no restrictions are imposed on the treatment policy m, then inequality (7) provides the sharp bound on $P(y_m \epsilon B \mid x)$. One may sometimes have prior information that, when combined with empirical knowledge of $[P(y_1 \mid x), P(y_0 \mid x)]$, makes the bound (7) computable. This section presents three cases.

4.1. INDEPENDENT OUTCOMES

Suppose it is known that the outcomes y_1 and y_0 are statistically independent, conditional on x. Then

$$(13)\quad P(y_1 \epsilon B \cap y_0 \epsilon B \mid x) = P(y_1 \epsilon B \mid x) P(y_0 \epsilon B \mid x).$$

Our second proposition follows immediately:

Proposition 2: Let $P(y_1 \mid x)$ and $P(y_0 \mid x)$ be known. Let it be known that y_1 and y_0 are statistically independent, conditional on x. Then

$$(14)\quad P(y_1 \epsilon B \mid x) P(y_0 \epsilon B \mid x) \leq P(y_m \epsilon B \mid x)$$
$$\leq P(y_1 \epsilon B \mid x) + P(y_0 \epsilon B \mid x) - P(y_1 \epsilon B \mid x) P(y_0 \epsilon B \mid x). \quad \blacksquare$$

Whereas the bound obtained in Proposition 1 was generically one-sided, the present bound is generically two-sided. The new lower bound on $P(y_m \mid x)$ is informative whenever $P(y_1 \epsilon B \mid x) > 0$ and $P(y_0 \epsilon B \mid x) > 0$. The upper bound is informative whenever $P(y_1 \epsilon B \mid x) < 1$ and $P(y_0 \epsilon B \mid x) < 1$.

Suppose that Y is the real line. By Proposition 2,

$$(15)\quad P(y_1 \leq u \mid x)P(y_0 \leq u \mid x) \;\leq\; P(y_m \leq u \mid x)$$
$$\leq\; P(y_1 \leq u \mid x) + P(y_0 \leq u \mid x) - P(y_1 \leq u \mid x)P(y_0 \leq u \mid x)$$

for all $u \in R^1$. Corollary 2.1 inverts (15) to obtain sharp bounds on quantiles of $P(y_m \mid x)$. The proof uses the same argument as was applied to prove Corollary 1.1, and so is omitted.

Corollary 2.1: Let $P(y_1 \mid x)$ and $P(y_0 \mid x)$ be known. Let it be known that y_1 and y_0 are statistically independent, conditional on x. Let Y be the real line. Let

$$r_2(\alpha \mid x) \equiv \inf_u \text{ s.t. } P(y_1 \leq u \mid x) + P(y_0 \leq u \mid x) - P(y_1 \leq u \mid x)P(y_0 \leq u \mid x) \geq \alpha$$

$$s_2(\alpha \mid x) \equiv \inf_u \text{ s.t. } P(y_1 \leq u \mid x)P(y_0 \leq u \mid x) \geq \alpha$$

Then

$$(16)\quad r_2(\alpha \mid x) \;\leq\; q_m(\alpha \mid x) \;\leq\; s_2(\alpha \mid x). \quad \blacksquare$$

4.2. SHIFTED OUTCOMES

Evaluation studies often assume that y_1 and y_0 are not only statistically dependent but functionally dependent. It is especially common to assume that real-valued outcomes are shifted versions of one another; that is,[8]

$$(17)\quad P(y_1 = y_0 + \delta \mid x) = 1,$$

for some $\delta \epsilon R^1$. Shifted outcomes is widely thought of as a convenient, relatively innocuous assumption. However, recent analyses of the selection problem make clear that it is a quite restrictive condition with strong identifying power. See Heckman and Robb (1985), Bjorklund and Moffitt (1987), Robinson (1989), Moffitt (1990), and Manski (1993).

Suppose that (17) holds. Knowledge of $P(y_1 \mid x)$ and $P(y_0 \mid x)$ implies knowledge of δ. So the joint distribution $P(y_1,y_0 \mid x)$ is known and the bound (7) is computable. Thus we have

Proposition 3: Let $P(y_1 \mid x)$ and $P(y_0 \mid x)$ be known. Let Y be the real line. Let it be known that $P(y_1=y_0+\delta \mid x) = 1$, for some $\delta \epsilon R^1$. Then δ is identified and

$$(18)\quad P[(y_0+\delta)\epsilon B \cap y_0\epsilon B \mid x] \le P(y_m\epsilon B \mid x)$$
$$\le P[(y_0+\delta)\epsilon B \mid x] + P(y_0\epsilon B \mid x) - P[(y_0+\delta)\epsilon B \cap y_0\epsilon B \mid x]. \quad \blacksquare$$

When $B = (-\infty,u]$, this bound takes a very simple form. Assume, without loss of generality, that $\delta \ge 0$. Then (18) becomes

$$(19)\quad P(y_0 \le u-\delta \mid x) \le P(y_m \le u \mid x) \le P(y_0 \le u \mid x)$$

or, equivalently,

$$(19')\quad P(y_1 \le u \mid x) \le P(y_m \le u \mid x) \le P(y_0 \le u \mid x).$$

Corollary 3.1 inverts (19') to obtain sharp bounds on quantiles of $P(y_m \mid x)$.

Corollary 3.1: Let $P(y_1 \mid x)$ and $P(y_0 \mid x)$ be known. Let Y be the real line. Let it be known that $P(y_1 = y_0 + \delta \mid x) = 1$, for some $\delta \geq 0$. Let

$$r_3(\alpha \mid x) \equiv \inf_u \text{ s.t. } P(y_0 \leq u \mid x) \geq \alpha$$

$$s_3(\alpha \mid x) \equiv \inf_u \text{ s.t. } P(y_1 \leq u \mid x) \geq \alpha.$$

Then

$$(20) \quad r_3(\alpha \mid x) \leq q_m(\alpha \mid x) \leq s_3(\alpha \mid x). \quad \blacksquare$$

4.3. ORDERED OUTCOMES

Outcomes y_1 and y_0 are said to be ordered with respect to a given set B if y_0 almost always falls in B when y_1 does; that is,[9]

$$(21) \quad P(y_0 \epsilon B \mid x, y_1 \epsilon B) = 1.$$

For example, let the outcomes be binary, taking the value 0 or 1. If $P(y_0 = 0 \mid x, y_1 = 0) = 1$, then the outcomes are ordered with respect to the set $B = \{0\}$. As another example, suppose that the outcomes are real-valued and that

(22) $P(y_1 \geq y_0 \mid x) = 1.$

Then y_1 and y_0 are ordered with respect to the sets $B = (-\infty, u]$, as $y_1 \leq u \Rightarrow y_0 \leq u$.

The assumption of ordered outcomes has earlier been discussed in the context of the Perry Preschool project. One may believe that receiving the intervention cannot possibly diminish a child's prospects for graduation. If so, then any child who receives the intervention and does not graduate would not graduate in the absence of the intervention. That is, $P(y_0=0 \mid x, y_1=0) = 1$.

If y_1 and y_0 are ordered with respect to B, then

(23) $P(y_1 \epsilon B \cap y_0 \epsilon B \mid x) = P(y_1 \epsilon B \mid x)$,

so the bound (7) is computable. In particular, we have

Proposition 4: Let $P(y_1 \mid x)$ and $P(y_0 \mid x)$ be known. Let it be known that $P(y_0 \epsilon B \mid x, y_1 \epsilon B) = 1$. Then

(24) $P(y_1 \epsilon B \mid x) \leq P(y_m \epsilon B \mid x) \leq P(y_0 \epsilon B \mid x)$. ■

An interesting result emerges when (24) is applied to real-valued outcomes satisfying (22). Letting $B = (-\infty, u]$, we find that (24) coincides with the bound (19') that holds when outcomes are known to be shifted. Thus it turns out that assumptions (17) and (22) have the same identifying power in the context of the mixing problem. Manski (1993) shows that these assumptions have different implications in the context of the selection problem.

5. Restrictions on the Treatment Policy

Sections 5.1 and 5.2 examine the restrictions on $P(y_m \mid x)$ implied by a set of polar treatment policies, in the absence of prior information about the outcome distribution. Section 5.1 supposes that treatment is statistically independent of outcomes, as in random assignment policies. Section 5.2 supposes that treatment minimizes or maximizes the probability that the realized outcome y_m falls in specified sets B, as in competing risks models and in the Roy model. Section 5.3 examines the quite different problem of inference when the fraction of the population receiving each treatment is known, but nothing is known about the composition of the subpopulations receiving each treatment.

5.1. TREATMENT INDEPENDENT OF OUTCOMES

Suppose it is known that the treatment z_m received by each person with covariates x is statistically independent of the person's outcomes (y_1,y_0). That is,

(25) $P[(y_1,y_0) \mid x,z_m] = P[(y_1,y_0) \mid x]$.

Then equation (2) reduces to

(26) $P(y_m \mid x) \;=\; P(y_1 \mid x)P(z_m=1 \mid x) + P(y_0 \mid x)P(z_m=0 \mid x).$

If the fraction $P(z_m \mid x)$ of the population receiving each treatment is known, then $P(y_m \mid x)$ is identified. Our present concern, however, is with the situation in which (25) is the only prior information available. In this case, the only restriction on the treatment distribution is that $P(z_m=1 \mid x)$ and $P(z_m=0 \mid x)$ must lie in the unit interval and add up to one. Hence Proposition 5 follows immediately:

<u>Proposition 5:</u> Let $P(y_1 \mid x)$ and $P(y_0 \mid x)$ be known. Let it be known that z_m is statistically independent of (y_1,y_0), conditional on x. Then

(27) $\min[P(y_1 \in B \mid x),P(y_0 \in B \mid x)] \;\le\; P(y_m \in B \mid x) \;\le\; \max[P(y_1 \in B \mid x),P(y_0 \in B \mid x)].$ ■

The bound of Proposition 5 is a subset of each of the bounds reported in Propositions 2 through 4, which left the treatment policy unspecified and imposed restrictions on the outcome distribution. This fact has a simple explanation. Equation (26) shows that, if z_m is statistically independent of (y_1,y_0), then $P(y_m \mid x)$ depends on the distribution of (y_1,y_0) only through the two marginal distributions $P(y_1 \mid x)$ and $P(y_0 \mid x)$. Hence, if one knows that z_m is independent of (y_1,y_0), then restrictions on the distribution of (y_1,y_0) have no identifying power.

Let Y be the real line and let $B = (-\infty,u]$. Inverting the bound in Proposition 5 produces the following bound on quantiles of $P(y_m \mid x)$:

Corollary 5.1: Let $P(y_1 \mid x)$ and $P(y_0 \mid x)$ be known. Let it be known that z_m is statistically independent of (y_1, y_0), conditional on x. Let Y be the real line. Let

$$r_5(\alpha \mid x) \equiv \inf_u \text{ s.t. } \max[P(y_1 \le u \mid x), P(y_0 \le u \mid x)] \ge \alpha.$$

$$s_5(\alpha \mid x) \equiv \inf_u \text{ s.t. } \min[P(y_1 \le u \mid x), P(y_0 \le u \mid x)] \ge \alpha.$$

Then

$$(28) \quad r_5(\alpha \mid x) \le q_m(\alpha \mid x) \le s_5(\alpha \mid x). \quad \blacksquare$$

5.2. OPTIMIZING TREATMENTS

To prove Proposition 1, we constructed two extreme treatment policies, one minimizing $P(y_m \epsilon B \mid x)$ and the other maximizing it. The former policy satisfies equation (5), while the latter satisfies (6). Suppose that one of these optimizing policies is actually implemented. What can be learned about $P(y_m \epsilon B \mid x)$ in the absence of prior restrictions on the outcome distribution?

The proof to Proposition 1 showed that the treatment policy minimizing $P(y_m \epsilon B \mid x)$ makes $P(y_m \epsilon B \mid x) = P(y_1 \epsilon B \cap y_0 \epsilon B \mid x)$, while the policy maximizing $P(y_m \epsilon B \mid x)$ makes $P(y_m \epsilon B \mid x) = P(y_1 \epsilon B \cup y_0 \epsilon B \mid x)$. Applying the Frechet Bound (8) yields Proposition 6:

Proposition 6: Let $P(y_1 \mid x)$ and $P(y_0 \mid x)$ be known.

A. Let it be known that the treatment policy satisfies (5). Then

$$(29)\quad \max[0, P(y_1 \in B \mid x) + P(y_0 \in B \mid x) - 1] \leq P(y_m \in B \mid x) \leq \min[P(y_1 \in B \mid x), P(y_0 \in B \mid x)].$$

B. Let it be known that the treatment policy satisfies (6). Then

$$(30)\quad \max[P(y_1 \in B \mid x), P(y_0 \in B \mid x)] \leq P(y_m \in B \mid x) \leq \min[P(y_1 \in B \mid x) + P(y_0 \in B \mid x), 1]. \quad \blacksquare$$

It is interesting to compare these bounds with those under other assumptions. In Part A, the lower bound coincides with the lower bound in the absence of prior information (see Proposition 1), while the upper bound coincides with the lower bound under the assumption that treatment is independent of the outcomes (see Proposition 5). In Part B, the lower bound coincides with the upper bound under the assumption that treatment is independent of the outcomes (see Proposition 5), while the upper bound coincides with the upper bound in the absence of prior information (see Proposition 1). Thus the three treatment policies examined in Propositions 5 and 6 imply that $P(y_m \in B \mid x)$ lies in mutually exclusive intervals, and these three intervals partition the range of values that is feasible in the absence of prior information.

SELECTION OF THE TREATMENT WITH THE LARGER OR SMALLER OUTCOME: Proposition 6 has important applications in economics and in survival analysis. Economic analyses of voluntary treatment policies often assume that Y is the real line and that the treatment yielding the larger outcome is selected, so

$$(31)\quad y_m = \max(y_1, y_0).$$

In the labor-economics literature on occupation choice, this is often called the Roy model (see Heckman and Honore, 1990). For any $u \in R^1$, treatment policy (31) makes $P(y_m \le u \mid x) = P(y_1 \le u \cap y_0 \le u \mid x)$. So this policy minimizes $P(y_m \le u \mid x)$. We may therefore apply part A of Proposition 6 to show that

$$(32)\quad \max[0, P(y_1 \le u \mid x) + P(y_0 \le u \mid x) - 1] \le P(y_m \le u \mid x)$$
$$\le \min[P(y_1 \le u \mid x), P(y_0 \le u \mid x)].$$

The *competing risks* model of survival analysis (see Kalbfleisch and Prentice, 1980) assumes that Y is the real line and that the treatment yielding the smaller outcome is selected, so

$$(33)\quad y_m = \min(y_1, y_0).$$

For any u, this treatment policy maximizes $P(y_m \le u \mid x)$. So Part B of Proposition 6 shows that

$$(34)\quad \max[P(y_1 \le u \mid x), P(y_0 \le u \mid x)] \le P(y_m \le u \mid x) \le \min[P(y_1 \le u \mid x) + P(y_0 \le u \mid x), 1].$$

Corollary 6.1 inverts the bounds (32) and (34) to produce bounds on quantiles of

$P(y_m \mid x)$.

Corollary 6.1: Let $P(y_1 \mid x)$ and $P(y_0 \mid x)$ be known. Let Y be the real line.

A. Let it be known that $y_m = \max(y_1, y_0)$. Then

$$(35)\quad s_5(\alpha \mid x) \le q_m(\alpha \mid x) \le s_1(\alpha \mid x).$$

B. Let it be known that $y_m = \min(y_1, y_0)$. Then

$$(36)\quad r_1(\alpha \mid x) \le q_m(\alpha \mid x) \le r_5(\alpha \mid x). \quad \blacksquare$$

TREATMENT A KNOWN FUNCTION OF THE OUTCOMES: The treatment policies selecting the treatment with the larger or smaller outcome are members of a class of policies making the treatment z_m received by each person a function of the person's outcomes (y_1, y_0). Suppose that the function $z_m(.,.)$: $YxY \rightarrow \{0,1\}$ mapping outcomes into treatments is known. Then the realized outcome y_m is a known function of (y_1, y_0), namely

$$(37)\quad y_m = y_1 z_m(y_1, y_0) + y_0[1 - z_m(y_1, y_0)].$$

With y_m a known function of (y_1, y_0), $P(y_m \mid x)$ is identified if information identifying the outcome distribution is available. In particular, the analysis of Section 4 implies that

$P(y_m \epsilon B \mid x)$ is identified if (y_1, y_0) are known to be statistically independent, shifted, or ordered outcomes.

5.3. KNOWN TREATMENT DISTRIBUTION

The restrictions on treatment policy examined in Sections 5.1 and 5.2 specify the rule used to make treatment assignments, but do not a priori constrain the fraction of the population receiving each treatment. It is also of interest to consider the reverse situation, where one knows the fraction receiving each treatment but does not know the rule used to make treatment assignments. For example, we noted earlier that resource constraints could limit implementation of the Perry Preschool treatment to part of the eligible population. Knowledge of the budget constraint and the cost of pre-schooling would suffice to determine the fraction of the population receiving the treatment. It may be more difficult to learn how school officials, social workers, and parents interact to determine which children receive the treatment.

Thus suppose that under policy m, a known fraction p of the persons with covariates x receive treatment y_0, the remaining fraction (1-p) receiving treatment y_1. So

$$(38) \quad P(z_m=0 \mid x) \; = \; p,$$

where p is known. No information is available on the rule used to make treatment assignments that satisfy (38).

Given (38), $P(y_m \mid x)$ may be written

(39) $P(y_m \mid x) = P(y_1 \mid x,z_m=1)(1-p) + P(y_0 \mid x,z_m=0)p$.

The distributions $[P(y_1 \mid x),P(y_0 \mid x)]$ may be written

(40a) $P(y_1 \mid x) = P(y_1 \mid x,z_m=1)(1-p) + P(y_1 \mid x,z_m=0)p$

and

(40b) $P(y_0 \mid x) = P(y_0 \mid x,z_m=1)(1-p) + P(y_0 \mid x,z_m=0)p$.

Knowledge of $P(y_1 \mid x)$ and p restricts $P(y_1 \mid x,z_m=1)$ and $P(y_1 \mid x,z_m=0)$ to pairs of distributions that satisfy (40a); similarly, knowledge of $P(y_0 \mid x)$ and p restricts $P(y_0 \mid x,z_m=1)$ and $P(y_0 \mid x,z_m=0)$ to pairs of distributions that satisfy (40b). Examination of the feasible pairs shows that $P(y_1 \mid x,z_m=1)$ and $P(y_0 \mid x,z_m=0)$ must lie in the following sets of distributions:

(41a) $P(y_1 \mid x,z_m=1) \in \Psi_{11}(p) \equiv \Psi \cap [\{P(y_1 \mid x)-p\psi\}/(1-p): \psi \in \Psi]$

and

(41b) $P(y_0 \mid x,z_m=0) \in \Psi_{00}(p) \equiv \Psi \cap [\{P(y_0 \mid x) - (1-p)\psi\}/p: \psi \in \Psi]$,

where Ψ denotes the set of all distributions on Y. It follows that $P(y_m \mid x)$ is a (1-p,p) mixture of a distribution in $\Psi_{11}(p)$ and one in $\Psi_{00}(p)$. That is,

$$(42) \quad P(y_m \mid x) \in [(1-p)\psi_{11} + p\psi_{00}: (\psi_{11},\psi_{00}) \in \Psi_{11}(p)\times\Psi_{00}(p)].$$

Relation (42) completely characterizes the restrictions on $P(y_m \mid x)$ implied by knowledge of $[P(y_1 \mid x),P(y_0 \mid x),P(z_m \mid x)]$, but the characterization is not transparent. Horowitz and Manski (1992) have analyzed the sets $\Psi_{11}(p)$ and $\Psi_{00}(p)$ in a recent study of the *contaminated sampling* problem, whose formal structure is similar to the problem studied here. Their Corollary 1.2 proves the following sharp bounds on $P(y_1\epsilon B \mid x,z_m=1)$ and $P(y_0\epsilon B \mid x,z_m=0)$:[10]

$$(43a) \quad \max[0,\{P(y_1\epsilon B \mid x)-p\}/(1-p)] \le P(y_1\epsilon B \mid x,z_m=1) \le \min[1,P(y_1\epsilon B \mid x)/(1-p)]$$

and

$$(43b) \quad \max[0,\{P(y_0\epsilon B \mid x)-(1-p)\}/p] \le P(y_0\epsilon B \mid x,z_m=0) \le \min[1,P(y_0\epsilon B \mid x)/p].$$

This and (38) imply Proposition 7:

Proposition 7: Let $P(y_1 \mid x)$ and $P(y_0 \mid x)$ be known. Let $P(z_m=0 \mid x) = p$, for known p. Then

$$(44) \quad \max[0,P(y_1\epsilon B \mid x)-p] + \max[0,P(y_0\epsilon B \mid x)-(1-p)] \le P(y_m\epsilon B \mid x)$$

$$\leq \min[1\text{-}p, P(y_1 \epsilon B \mid x)] + \min[p, P(y_0 \epsilon B \mid x)]. \blacksquare$$

Inverting this bound yields Corollary 7.1.

Corollary 7.1: Let $P(y_1 \mid x)$ and $P(y_0 \mid x)$ be known. Let $P(z_m=0 \mid x) = p$, for known p. Let Y be the real line. Let

$$r_{7p}(\alpha \mid x) \equiv \inf_u \text{ s.t. } \min[1\text{-}p, P(y_1 \leq u \mid x)] + \min[p, P(y_0 \leq u \mid x)] \geq \alpha$$

$$s_{7p}(\alpha \mid x) \equiv \inf_u \text{ s.t. } \max[0, P(y_1 \leq u \mid x)\text{-}p] + \max[0, P(y_0 \leq u \mid x)\text{-}(1\text{-}p)] \geq \alpha.$$

Then

$$(45) \quad r_{7p}(\alpha \mid x) \leq q_m(\alpha \mid x) \leq s_{7p}(\alpha \mid x). \blacksquare$$

EVIDENCE ON ONE TREATMENT: Throughout this paper, I have assumed that empirical evidence is available for both treatments. Suppose now that such evidence is available for only one treatment, say treatment 1; so $P(y_1 \mid x)$ is known and $P(y_0 \mid x)$ is unrestricted. In the absence of information on the fraction of the population receiving each treatment, nothing can be learned about $P(y_m \mid x)$. After all, $P(z_m=0 \mid x) = 1$ might hold, in which case $P(y_m \mid x) = P(y_0 \mid x)$. On the other hand, some inference on $P(y_m \mid x)$ is possible if $P(z_m \mid x)$ is known. Proposition 8 and Corollary 8.1 provide the results.

Proposition 8: Let $P(y_1 \mid x)$ be known. Let $P(z_m=0 \mid x) = p$, for known p. Then

(46) $\max[0, P(y_1 \epsilon B \mid x)-p] \leq P(y_m \epsilon B \mid x) \leq \min[1, P(y_1 \epsilon B \mid x)+p]$. ∎

PROOF: With $P(y_1 \mid x)$ and p known, (43a) continues to give the sharp bound on $P(y_1 \epsilon B \mid x, z_m=1)$. With $P(y_0 \mid x)$ unknown, the bound (43b) on $P(y_0 \epsilon B \mid x, z_m=0)$ is no longer available; all we know is that $0 \leq P(y_0 \epsilon B \mid x, z_m=0) \leq 1$. This and (38) imply (46).

Q.E.D.

Corollary 8.1: Let $P(y_1 \mid x)$ be known. Let $P(z_m=0 \mid x) = p$, for known p. Let Y be the real line. Let

$$r_{8p}(\alpha \mid x) \equiv \inf_u \text{ s.t. } \min[1, P(y_1 \leq u \mid x)+p] \geq \alpha$$

$$s_{8p}(\alpha \mid x) \equiv \inf_u \text{ s.t. } \max[0, P(y_1 \leq u \mid x)-p] \geq \alpha.$$

Then

(47) $r_{8p}(\alpha \mid x) \leq q_m(\alpha \mid x) \leq s_{8p}(\alpha \mid x)$. ∎

The lower bound on $P(y_m \epsilon B \mid x)$ is informative if $P(y_1 \epsilon B \mid x) > p$; the upper if $P(y_1 \epsilon B \mid x) < 1 - p$. When the bound is informative above and below, it restricts $P(y_m \epsilon B \mid x)$ to an interval of width 2p, centered at $P(y_1 \epsilon B \mid x)$. In contrast to the case when empirical evidence is available for both treatments, the present bounds on quantiles are not always informative. The lower bound is informative if $p < \alpha$; the upper if $p < 1 - \alpha$.

6. Identifying Combinations of Assumptions

Propositions 1 through 8 assume enough empirical evidence and prior information to bound event probabilities $P(y_m \epsilon B \mid x)$, but not enough to identify them. In Section 5, we noted in passing some assumptions that do suffice to identify $P(y_m \epsilon B \mid x)$. Proposition 9 presents these simple findings formally.

Proposition 9: Let $P(y_1 \mid x)$ and $P(y_0 \mid x)$ be known.

A. Let it be known that z_m is statistically independent of (y_1, y_0), conditional on x. Let $P(z_m = 0 \mid x) = p$, for known p. Then

$$(48)\quad P(y_m \epsilon B \mid x) \;=\; P(y_1 \epsilon B \mid x)(1-p) + P(y_0 \epsilon B \mid x)p$$

is identified.

B. Let $z_m = z_m(y_1, y_0)$ for some known function $z_m(\cdot,\cdot)$: $Y \times Y \rightarrow \{0,1\}$. Let it be known that y_1 and y_0 are either statistically independent, shifted, or ordered outcomes, conditional on x. Then

$$(49)\quad P(y_m \epsilon B \mid x) \;=\; P[y_1 z_m(y_1, y_0) + y_0\{1 - z_m(y_1, y_0)\} \epsilon B \mid x]$$

is identified. ■

Notes

1. In practice there often are multiple feasible treatments, but this paper restricts attention to the two-treatment case assumed in most of the literature. In the literature on experimentation, it is common to call one of these the *treatment* or *intervention*, and the other the *control*.

2. Of course one might observe realizations under more than one policy. Work on selection problems has focussed on the case in which only one policy is observed.

3. The mixing problem should not be confused with the converse problem: What does knowledge of $P[y_1 z_m + y_0(1-z_m) \mid x]$ imply about $[P(y_1 \mid x), P(y_0 \mid x), P(z_m \mid x)]$? The latter is sometimes referred to as a *mixture* problem.

4. A more demanding technical challenge, not addressed here, is to determine the identifiability of the conditional mean $E(y_m \mid x)$.

5. Prior information restricting the marginal distributions $P(y_1 \mid x)$ and $P(y_0 \mid x)$ has no identifying power as these distributions are identified by the empirical evidence. Such restrictions may improve the precision of sample estimates of $P(y_1 \mid x)$ and $P(y_0 \mid x)$, but this usage is distinct from the identification concerns of the present paper.

6. The estimates of $P(y_1 = 1 \mid x)$ and $P(y_0 = 1 \mid x)$ are based on the 58 treatment-group members and 63 control-group members from whom the investigators obtained graduation data.

7. See Ord (1972) for a brief exposition of the Frechet bounds, and Ruschendorf (1981) for a rather general analysis.

8. Knowledge of the marginal distributions $P(y_1 \mid x)$ and $P(y_0 \mid x)$ makes the shifted-outcome assumption a testable hypothesis. If (17) holds, $P(y_1 \mid x)$ and $P(y_0 \mid x)$ must be the same up to a translation of location. In contrast, the statistical independence assumption of Section 4.1 is not testable, as it implies no restrictions on $P(y_1 \mid x)$ and $P(y_0 \mid x)$.

9. Knowledge of the marginal distributions $P(y_1 \mid x)$ and $P(y_0 \mid x)$ makes the ordered-outcomes assumption a testable hypothesis. If (21) holds, then $P(y_0 \epsilon B \mid x) \geq P(y_1 \epsilon B \mid x)$.

10. That (43a) and (43b) are bounds follows immediately from (41a) and (41b) respectively. It is a bit more work to show that these bounds are sharp.

References

Berrueta-Clement, J., Schweinhart, L., Barnett, W., Epstein, A., and D. Weikart(1984), Changed Lives: The Effects of the Perry Preschool Program on Youths Through Age 19, Ypsilanti, Michigan: High/Scope Press.

Bjorklund, A. and R. Moffitt (1987), "Estimation of Wage Gains and Welfare Gains in Self-Selection Models," Review of Economics and Statistics, 69, 42-49.

Frechet, M.(1951), "Sur Les Tableaux de Correlation Donte les Marges sont Donnees", Annals de Universite de Lyon A, Series 3, 14, 53-77.

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.

Holden, C.(1990), "Head Start Enters Adulthood," Science, 247, 1400-1402.

Horowitz, J. and C. Manski(1992), "Identification and Robustness in the Presence of Errors in Data," Social Systems Research Institute Paper 9209, University of Wisconsin-Madison.

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

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

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

Manski, C.(1993), "The Selection Problem," in C. Sims (editor), Advances in Econometrics, Cambridge, UK: Cambridge University Press.

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

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

Moffitt, R.(1992), "Evaluation Methods for Program Entry Effects," in Manski, C. and I. Garfinkel (editors), Evaluating Welfare and Training Programs, Cambridge, Mass.: Harvard University Press.

Ord, J.(1972), Families of Frequency Distributions, Griffin's Statistical Monographs & Courses No.30, New York: Hafner.

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.

Ruschendorf, L.(1981), "Sharpness of Frechet-Bounds," Zeitschrift fur Wahrscheinlichkeitstheorie und verwandte Gebiete, 57, 293-302.