Nonparametric Bounds on Treatment Effects

By CHARLES F. MANSKI*

Assume that each member of a population is characterized by values for the variables (y_A, y_B, z, x). Here x is a vector describing a person and z is a binary variable indicating which of two treatments this person receives. The treatments are labelled A and B. The variables y_A and y_B are scalar measures of the outcomes of the two treatments.

For example, a cancer patient might be treated by (A) drug therapy or (B) surgery. The relevant outcome y might be life span following treatment. An unemployed worker might be given (A) vocational training or (B) job search assistance. Here the relevant outcome might be labor force status following treatment.

Assume that a random sample is drawn and that one observes the realizations of (z, x) and of the outcome under the treatment received. Thus y_A is observed if treatment A is received but is a latent variable if treatment B is received. Similarly, y_B is either observed or latent.

Suppose that one wants to learn the difference in expected outcome if all persons with attributes x were assigned to treatment A or B. This "treatment effect" is

$$(1)\quad t(x) \equiv E(y_B|x) - E(y_A|x)$$
$$= E(y_B|x, z=A)P(z=A|x)$$
$$+ E(y_B|x, z=B)P(z=B|x)$$
$$- E(y_A|x, z=A)P(z=A|x)$$
$$- E(y_A|x, z=B)P(z=B|x).$$

*Department of Economics and Institute for Research on Poverty, University of Wisconsin, Madison, WI 53706. This research is supported by NSF grant no. SES-8808276 and by a grant from the U.S. Department of Health and Human Services to the Institute for Research on Poverty. I have benefited from the comments of Arthur Goldberger and Robert Moffitt.

The central problem is identification. The data are from a population in which some people described by x received treatment A and the rest received B. The sampling process identifies the expected outcomes under the treatment received, $E(y_A|x, z=A)$ and $E(y_B|x, z=B)$. It also identifies the treatment-selection probabilities $P(z|x)$. But, the sampling process does not identify $E(y_A|x, z=B)$ and $E(z_B|x, z=A)$. Hence it does not identify the treatment effect.

An extensive literature on the estimation of treatment effects brings to bear prior information that, in conjunction with the sampling process, does identify $t(x)$. (See, for example, G. S. Maddala, 1983, and James Heckman and Richard Robb, 1985.) Two approaches have been dominant. One assumes that, conditional on x, y_A and y_B are mean-independent of z (i.e., $E(y_A|x, z=A) = E(y_A|x, z=B)$ and $E(y_B|x, z=A) = E(y_B|x, z=B)$). This assumption, routinely invoked in experiments with random assignment to treatment, implies that

$$(2)\quad t(x) = E(y_B|x, z=B) - E(y_A|x, z=A).$$

The second approach imposes identifying restrictions through a latent-variable model explaining treatment selections and outcomes. The latent-variable-model approach is widely used in settings where people self-select into treatment.

Suppose that one cannot confidently assert mean-independence, an identifying latent-variable model, or any other restriction that identifies the treatment effect. It might then seem that useful inference is impossible. This paper proves otherwise. Section I applies results from my earlier paper (1989) to show that an informative bound on the treatment effect holds whenever the outcomes y_A and y_B are themselves bounded. Section II derives a bound applicable when the treatment selection rule is to choose the treatment with the better outcome. Section III shows that the bounds of Sections I and II can be tightened if some component of x affects treatment selection but not treatment outcome. Section IV briefly discusses estimation of the bounds.

I. Bounded Outcomes

Suppose that, conditional on x, y_A is bounded within some known interval $[K_{A0x}, K_{A1x}]$, where $-\infty < K_{A0x} \leq K_{A1x} < \infty$. Then obviously $E(y_A|x) \in [K_{A0x}, K_{A1x}]$. My earlier paper observes that the sampling process identifies a tighter bound, namely,

$$(3) \quad E(y_A|x) \in Y_A(x) \equiv [E(y_A|x, z=A)P(z=A|x) + K_{A0x}P(z=B|x), \; E(y_A|x, z=A)P(z=A|x) + K_{A1x}P(z=B|x)].$$

The lower bound is the value $E(y_A|x)$ takes if y_A equals its lower bound for all those who receive treatment B; the upper bound is determined similarly.

The width of the bound $Y_A(x)$ is $(K_{A1x} - K_{A0x})P(z=B|x)$. So the bound is informative if $P(z=B|x)<1$. The bound is operational because $E(y_A|x, z=A)$ and $P(z|x)$ are identified by the sampling process. In practice, one can estimate $E(y_A|x, z=A)$ and $P(z|x)$ nonparametrically, yielding a nonparametric estimate for the bound (see Section IV).

Now suppose that y_B is also bounded, the interval being $[K_{B0x}, K_{B1x}]$. Then the treatment effect must lie in the interval $[K_{B0x} - K_{A1x}, K_{B1x} - K_{A0x}]$. The sampling process identifies a tighter bound, obtained by applying (3) and the analogous bound for $E(y_B|x)$ to (1). The result is

$$(4) \quad t(x) \in T(x) \equiv [K_{B0x}P(z=A|x) + E(y_B|x, z=B)P(z=B|x) - E(y_A|x, z=A)P(z=A|x) - K_{A1x}P(z=B|x), \; K_{B1x}P(z=A|x) + E(y_B|x, z=B)P(z=B|x) - E(y_A|x, z=A)P(z=A|x) - K_{A0x}P(z=B|x)].$$

The lower bound on $t(x)$ is the difference between the lower bound on $E(y_B|x)$ and the upper bound on $E(y_A|x)$. The upper bound on $t(x)$ is determined similarly.

The width of the bound $T(x)$ is

$$(5) \quad w(x) \equiv (K_{B1x} - K_{B0x})P(z=A|x) + (K_{A1x} - K_{A0x})P(z=B|x).$$

In general, this width depends on the treatment-selection probabilities $P(z|x)$. Suppose, however, that the bounds on y_A and y_B are the same (i.e., $[K_{A0x}, K_{A1x}] = [K_{B0x}, K_{B1x}]$). Then,

$$(6) \quad w(x) = K_{1x} - K_{0x},$$

where $[K_{0x}, K_{1x}]$ is the common bound on the outcomes. The bound available without the sample data is $t(x) \in [K_{0x} - K_{1x}, K_{1x} - K_{0x}]$. Thus, when the bounds on y_A and y_B are the same, exploitation of the sampling process allows one to bound $t(x)$ to one-half of its otherwise possible range. In this case, the bound necessarily covers zero; it cannot identify the sign of the treatment effect.

The remainder of this section describes a class of applications in which the bound $T(x)$ is particularly useful—binary logical outcomes.

In many applications, the treatment outcome is a logical yes/no 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 vocational training program may be (completed $=1$, not completed $=0$). In both cases, $K_{A0x}=K_{B0x}=0$ and $K_{A1x}=K_{B1x}=1$, so the treatment effect must lie in the interval $[-1,1]$. The expected value of a one/zero indicator is the probability that the indicator equals one. So the bound $T(x)$ reduces to

$$(7)\quad T(x) = [P(y_B=1|x,z=B)P(z=B|x) - P(y_A=1|x,z=A)P(z=A|x) - P(z=B|x), P(z=A|x) + P(y_B=1|x,z=B)P(z=B|x) - P(y_A=1|x,z=A)P(z=A|x)].$$

The bound width is $w(x)=1$.

Binary logical variables are bounded by definition rather than by assumption. So we find that the sampling process alone, unaccompanied by prior information, suffices to bound the treatment effect to one-half its otherwise possible range.

II. Selection of The Treatment with the Better Outcome

In some settings the treatment-selection rule has the form

$$(8)\qquad z=B \Leftrightarrow y_B \geq y_A.$$

For example, a doctor may prescribe the more effective of two medical treatments. An unemployed worker may choose the employment program with the higher return.

If (8) holds, the bound $T(x)$ obtained in Section I can be tightened. By (8),

$$(9)\quad \begin{aligned} E(y_A|x,z=B) &= E(y_A|x,y_A\leq y_B) \\ &\leq E(y_A|x,y_A>y_B) \\ &= E(y_A|x,z=A) \\ E(y_B|x,z=A) &= E(y_B|x,y_B<y_A) \\ &\leq E(y_B|x,y_B\geq y_A) \\ &= E(y_B|x,z=B). \end{aligned}$$

Thus $E(y_A|x,z=A)$ and $E(y_B|x,z=B)$ are upper bounds on $E(y_A|x,z=B)$ and $E(y_B|x,z=A)$, respectively. The conditions $E(y_A|x,z=A)\leq K_{A1x}$ and $E(y_B|x,z=B)\leq K_{B1x}$ must hold. Hence knowing that (8) holds permits one to tighten the bound (4) on the treatment effect to

$$(10)\quad T(x) = [K_{B0x}P(z=A|x) + E(y_B|x,z=B)P(z=B|x) - E(y_A|x,z=A), E(y_B|x,z=B) - E(y_A|x,z=A)P(z=A|x) - K_{A0x}P(z=B|x)].$$

The tightened bound may or may not lie entirely to one side of zero. If it does, the sign of the treatment effect is identified.

III. Level-Set Restrictions

The bound $T(x)$ on the treatment effect at a given value of x does not constrain the treatment effect elsewhere. This is to be expected as no restrictions have been imposed on the behavior of $t(x)$ as a function of x. Suppose that one has information on the way $t(x)$, or its determinants, vary with x. Then one may be able to obtain a bound tighter than $T(x)$.

This section investigates the additional identifying power of level-set restrictions. A level-set restriction is an assertion that some function of x is constant on some $X_0 \subset X$, where X is the set of all possible values of x. An important special case is the exclusion restriction. Here one lets $x \equiv (x_1, x_2)$ and asserts that, holding x_1 fixed, a function of x does not vary with x_2. Thus the function is constant on the set $X_0 \equiv \{x_1\} \times X_2$, where $\{x_1\}$ is the set containing only the point x_1, and where X_2 is the set of all possible values of x_2.

A. *Level-Set Restrictions on the Treatment Effect*

It is often assumed in applications that the treatment effect does not vary with x. In particular, many studies specify a linear model with $E(y_A|x) = x\beta$ and $E(y_B|x) = x\beta + \alpha$, implying that $t(x) = \alpha$. The assumption that $t(x)$ is constant on all of X is a leading example of a level-set restriction.

Suppose it is known that $t(x)$ is constant on some set X_0. Then the collection of bounds $T(x)$, $x \in X_0$ must have a non-null intersection that contains the common value of the treatment effect. That is, for each $\xi \in X_0$,

$$(11)\quad t(\xi) \in T_0(X_0) \equiv \bigcap_{x \in X_0} T(x)$$
$$= \Big[\sup_{x \in X_0} \{ K_{B0x} P(z = A|x) + E(y_B|x, z = B) P(z = B|x) - E(y_A|x, z = A) P(z = A|x) - K_{A1x} P(z = B|x) \},$$
$$\inf_{x \in X_0} \{ K_{B1x} P(z = A|x) + E(y_B|x, z = B) P(z = B|x) - E(y_A|x, z = A) P(z = A|x) - K_{A0x} P(z = B|x) \} \Big].$$

The bound $T_0(X_0)$ improves on $T(\xi)$ for at least some ξ in X_0 unless $T(\cdot)$ is constant on X_0. Constancy of $T(\cdot)$ can occur in various ways. The one most likely to arise in practice is inclusion in x of an irrelevant component, one that affects neither K_{A0x}, K_{A1x}, K_{B0x}, K_{B1x}, $P(z|x)$, $E(y_A|x, z = A)$, nor $E(y_B|x, z = B)$. A restriction excluding this component from the treatment effect has no bite.

Although $T_0(X_0)$ improves on $T(x)$, it typically does not fix the sign of the treatment effect. If $T(x)$ covers zero for all $x \in X_0$, then so does $T_0(X_0)$. It was pointed out in Section I that $T(x)$ does cover zero whenever the bounds on y_A and y_B are the same.

B. *Level-Set Restrictions on the Outcome Regressions*

Suppose it is known that $E(y_A|x)$ is constant on some set $X_{A0} \subset X$, and that $E(y_B|x)$ is constant on some $X_{B0} \subset X$. (This includes cases in which one of the restrictions is trivial; the set X_{B0}, for example, might contain just one point.) Let $X_{AB0} \equiv X_{A0} \cap X_{B0}$. The reasoning used in Section A above implies that, for each $\xi \in X_{AB0}$, the bounds $Y_A(\xi)$ and $Y_B(\xi)$ defined in (3) can be tightened to

$$(12)\quad E(y_a|\xi) \in Y_{A0}(X_{A0}) \equiv \bigcap_{x \in X_{A0}} Y_A(x)$$
$$= \Big[\sup_{x \in X_{A0}} \{ E(y_A|x, z = A) P(z = A|x) + K_{A0x} P(z = B|x) \},$$
$$\inf_{x \in X_{A0}} \{ E(y_A|x, z = A) P(z = A|x) + K_{A1x} P(z = B|x) \} \Big]$$

$$E(y_B|\xi) \in Y_{B0}(X_{B0}) \equiv \bigcap_{x \in X_{B0}} Y_B(x)$$
$$= \Big[\sup_{x \in X_{B0}} \{ E(y_B|x, z = B) P(z = B|x) + K_{B0x} P(z = A|x) \},$$
$$\inf_{x \in X_{B0}} \{ E(y_B|x, z = B) P(z = B|x) + K_{B1x} P(z = A|x) \} \Big].$$

These bounds on $E(y_A|x)$ and $E(y_B|x)$ imply a bound on $t(\xi)$, namely,

$$(13)\quad t(\xi) \in T_{AB0}(X_{AB0})$$
$$\equiv \Bigg[\sup_{x \in X_{B0}} \{ E(y_B|x, z=B) P(z=B|x) + K_{B0x} P(z=A|x) \} - \inf_{x \in X_{A0}} \{ E(y_A|x, z=A) \times P(z=A|x) + K_{A1x} P(z=B|x) \},$$
$$\inf_{x \in X_{B0}} \{ E(y_B|x, z=B) P(z=B|x) + K_{B1x} P(z=A|x) \} - \sup_{x \in X_{A0}} \{ E(y_A|x, z=A) \times P(z=A|x) + K_{A0x} P(z=B|x) \} \Bigg].$$

The treatment effect is constant on X_{AB0}. Hence the bound $T_0(X_{AB0})$ also applies here. Comparison of (11) and (13) shows that $T_{AB0}(X_{AB0}) \subset T_0(X_{AB0})$. It is intuitive that the present bound should improve on the earlier one. The derivation of $T_0(X_{AB0})$ presumed only that $t(x)$ is constant on X_{AB0}. The derivation of $T_{AB0}(X_{AB0})$ imposed the stronger restriction that $E(y_A|x)$ is constant on X_{A0} and $E(y_B|x)$ is constant on X_{B0}.

The bound $T_{AB0}(X_{AB0})$ may lie entirely to one side of zero. If so, the sign of the treatment effect is identified.

IV. Estimation of the Bounds

The bounds developed in Sections I, II, and III are functions of $E(y_A|x, z=A)$, $E(y_B|x, z=B)$, and $P(z|x)$. These quantities are identified by the sampling process and so generally can be estimated consistently.

If the conditioning variable x takes finitely many values, estimation is classical. $E(y_A|x, z=A)$, $E(y_B|x, z=B)$, and $P(z|x)$ can be estimated by the corresponding sample averages. $T(x)$ can be estimated by inserting these averages into (4) or (10), as appropriate. $T_0(X_0)$, being a finite intersection of the $T(x)$, can be estimated by the intersection of the estimates of $T(x)$. Assuming that the level-set restriction on $t(x)$ is correct, this intersection is nonempty with probability approaching one as the sample size increases. The bound $T_{AB0}(X_{AB0})$ can be estimated similarly.

If x has continuous components, nonparametric regression methods may be applied to estimate $E(y_A|x, z=A)$, $E(y_B|x, z=B)$, and $P(z|x)$. My earlier paper exposits the main issues and presents an empirical illustration estimating the bound $Y_A(x)$ on $E(y_A|x)$. This work can be applied directly to estimate the bound $T(x)$. Estimation of $T_0(X_0)$ is a more subtle problem, because this bound is the intersection of the infinitely many bounds $T(x)$, $x \in X_0$. A plausible approach, that warrants study, is to estimate $T_0(X_{N0})$. Here N is the sample size and X_{N0}, $N=1,\dots,\infty$ is a sequence of finite subsets of X_0, chosen to converge to a set dense in X_0. The bound $T_{AB0}(X_{AB0})$ could be estimated similarly.

REFERENCES

Heckman, James and Robb, Richard, "Alternative Methods for Evaluating the Impact of Interventions," in J. Heckman and B. Singer, eds., *Longitudinal Analysis of Labor Market Data*, Cambridge: Cambridge University Press, 1985.

Manski, Charles F., "Anatomy of the Selection Problem," *Journal of Human Resources*, Summer 1989, *24*, 343–60.

Maddala, G. S., *Limited-Dependent and Qualitative Variables in Econometrics*, Cambridge: Cambridge University Press, 1983.