PRELIMINARY DRAFT: PLEASE DO NOT CITE WITHOUT PERMISSION

Strong Instrumental Variables

John V. Pepper
Department of Economics
University of Virginia
P.O. Box 400182
Charlottesville, VA 22904-4182
jvp3m@virginia.edu

Preliminary Draft

January 2002

Abstract:
This paper evaluates the identifying power of an exclusion restriction on the joint distribution of all response functions. Balke and Pearl (1997) show that this joint instrumental variable (IV) assumption implies a sharp bound that solves a complex linear programming problem. Using the same formal setup as in Manski (1990, 1994) and Manski and Pepper (2000), I show that the Balke and Pearl bounds have a much simpler presentation. Furthermore, in this presentation, the results apply for general instrumental variables and treatments, not just the binary variables examined by Balke and Pearl. The findings reported here add to the literature developing nonparametric bounds on treatment effects.

PRELIMINARY DRAFT: PLEASE DO NOT CITE WITHOUT PERMISSION

I. Introduction

A developing literature has examined the identifying power of instrumental variable assumptions in the absence of parametric restrictions on the form of response functions. Much of this literature has focused on the identifying power of different mean independence assumptions.[1] Manski (1990, 1994) and Robins (1989) show that mean independence assumptions imply sharp bounds on mean outcomes and average treatment effects. [2] The Manski-Robins bound, however, may not be sharp under the statistical independence that holds in classical randomized experiments, where response functions are jointly independent of assigned treatments, and other settings. In models with binary outcomes, treatments and instruments, Balke and Pearl (1997) show that the joint instrumental variable assumption implies a sharp bound that solves a complex linear programming problem. They find that the sharp bound improves on Manski-Robins bound for some data configurations but not for others. Unfortunately, the BP findings are difficult to understand and an intuitive interpretation of their finding has been elusive. This being the case, Robins and Greenland (1996) suggest use of the relatively simple albeit not necessarily sharp bound as a practical expedient.

This paper reevaluates the nonparametric identifying power of an exclusion restriction on the joint distribution of the response functions. Using the same formal setup as in Manski (1990, 1994) and Manski and Pepper (2000), I make two contributions to the literature. First, I show that the Balke and Pearl bounds have a much simpler presentation. Second, the resulting bounds apply for general instrumental variables and treatments, not just the binary

[1] There are several other contributions to this literature. Manski and Pepper (2000) evaluated the identify power of a mean monotonicity assumption. Hotz, Mullins, and Sanders (1997) have studied *contaminated instrument* assumptions which suppose that a mean-independence assumption holds in a population of interest, but the observed population is a probability mixture of the population of interest and one in which the assumption does not hold.

[2] Manski (1990) and Robins (1989) independently reported the same bound, the former assuming mean independence and the latter assuming statistical independence.

variables examined by Balke and Pearl. While I generalize the problem along some dimensions, the form of the sharp bound when outcomes are not binary remains entirely an open question.

This paper is organized in following way. In Section 2, I introduce the notation and the basic assumptions. In Section 3, I describe the Manski (1990) bounds under the mean independence assumption. Section 4 formalizes the strong instrumental variable bounds, shows that these bounds are sharp, and considers cases where the joint instrumental variable assumption is informative relative to a mean independence assumption. Section 5 briefly applies these bounds to the classical question of evaluating an experiment with noncompliance (not complete).

II. Model

This paper uses the same formal setup as in Manski (19978), Manski and Pepper (2000), and elsewhere. Each member j of population J has observable covariates $x_j \in X$ and a response function $y_j(\cdot): T \rightarrow Y$ mapping the mutually exclusive and exhaustive *treatments* $t \in T$ into *outcomes* $y_j(t) \in Y$. The outcomes Y are assumed to be binary. Person j has a realized treatment $z_j \in T$ and a realized outcome $y_j \in y_j(z_j)$, both of which are observable. The latent outcomes $y_j(t)$, $t \neq z_j$ are not observable. An empirical researcher learns the distribution P(x, z, y) of covariates, realized treatments, and realized outcomes by observing a random sample of the population. The researcher's problem is to combine this empirical evidence with assumptions in order to learn about the distribution $P[y(\cdot)]$ of response functions, or perhaps the conditional distributions $P[y(\cdot)| x]$.

With this background, I formally define the IV assumptions to be studied here. Let x = (w, v) and $X = W \times V$. Each value of (w, v) defines an observable sub-population of persons. The familiar mean-independence form of IV assumption is that, for each $t \in T$ and each value of w, the mean value of y(t) is the same in all of the sub-populations (w, v = u), $u \in V$. Thus, IV Assumption: Covariate v is an instrumental variable in the sense of mean- independence if, for each t 0 T, each value of w and v,

(1) P[y(t) = 1 | w, v] = P[y(t)= 1| w].

The strong instrumental variable assumption (SIV) is that the variable v is independent of the joint distribution of response functions. That is,

SIV Assumption: Covariate v is an strong instrumental variable in the sense of statistical independence if, for all t 0 T, each value of w and v,

(2) P[y(1), y(2), … , y(T)| w, v] = P[y(1), y(2), … , y(T)| w].

My interest is evaluating what can be learned about the mean response functions, P[y(t) = 1 | w], given the SIV assumption alone. To simply the exposition, I henceforth leave implicit the conditioning on w maintained in the definitions of instrumental variable assumptions. To keep the focus on identification, I treat identified quantities as known.

III. The IV Bound

The starting point for determination of the identifying power of the IV assumption in (1) is the no-assumptions bound on P[y(t)=1] reported in Manski (1989). Use the law of total probability to write

$$(3) \qquad P[y(t) = 1] = P[y(t) \mid z = t] \cdot P(z = t) + P[y(t) = 1 \mid z \neq t] \cdot P(z \neq t).$$

The sampling process identifies each of the quantities on the right side except for the censored mean $P[y(t) = 1 \mid z \neq t]$. In the absence of assumptions, all that is known about this censored mean is that lies between 0 and 1. This implies the sharp bound

$$(4) \qquad P[Y(1) = 1 \mid Z=1]\, P[Z=1] \leq P[Y(1) = 1] \leq P[Y(1) = 1 \mid Z=1]\, P[Z=1] + P(z \neq t).$$

Under the IV assumption in (1), $P[y(t) = 1|v]$ is constant across $v \in V$. Thus, it follows that the common value $P[y(t) = 1]$ lies in the intersection of the bounds (4) across the elements of V. Any point in this intersection is feasible. Thus, for all $v \in V$, we obtain the sharp bound

$$(5) \qquad \sup_{v \in V} \{P[y(t)=1 \mid v, z = t] \cdot P(z = t \mid v)\} \leq P[y(t)=1] \leq \inf_{v \in V} \{P[y(t)=1 \mid v, z = t] \cdot P(z = t \mid v) + P(z \neq t \mid v)\}.$$

IV. The SIV Bound

Similar methods result in bounds on the joint distribution and the resulting marginal distribution. With binary treatments, outcomes and instruments, Balke and Pearl (1997) find sharp bounds on the marginal distributions that result from the solution to complex linear programming problem. We begin in Section 4.1 by considering the same case, except that we relax the restriction that instruments are binary. In this more general setting, I replicate the Balke and Pearl bounds using a much simpler algorithm, and in Section 4.2 illustrates the power of the SIV assumption. In Section 4.3, I extend these results to apply to multiple treatments.

IV.1 Binary Treatments

The starting point for determination of the identifying power of the SIV assumption in (2) is the no-assumptions bound on the joint distribution, P[y(0), y(1)]. Consider, for instance, P[Y(1) = 1, Y(0) = 1]. Using the law of total probability to write

(6) P[Y(1) = 1,Y(0) = 1] = P[Y(1) = 1,Y(0) = 1 | Z = 1] P[Z = 1] + P[Y(1) = 1,Y(0) = 1 | Z = 0] P[Z = 0].

Since y is bounded in the unit interval it follows from Frechet (1951) that the sharp bound on the joint distribution is

(7) 0 # P[Y(1) = 1,Y(0) = 1] # P[Y(1) = 1 | Z = 1] P[Z = 1] + P[Y(0) = 1 | Z = 0] P[Z = 0].

Notice that the bounds on the joint distribution do not sum to the sharp bounds on the marginal distribution. After all, if P[Y(1) = 1,Y(0) = 1 | Z = 1] = P[Y(1) = 1 | Z = 1] then P[Y(1) = 1,Y(0) = 0 | Z = 1] = 0.

Still, the SIV assumption in (2) may be informative. Under the SIV assumption, P[y(0), y(1) |v] is constant across v 0 V. Thus, it follows that the common value P[y(0), y(1)] lies in the intersection of the bounds (7) across the elements of V. Any point in this intersection is feasible. Thus, for all v 0 V, we obtain the sharp bound

(8) $P[\ Y(1) = 1, Y(0) = 1\ ]\ \#$

$\inf \{ P[\ Y(1) = 1 \mid Z = 1, v]\ P[\ Z = 1 \mid v] + P[Y(0) = 1 \mid Z = 0, v]\ P[\ Z = 0 \mid v] \}.$
v 0 V

Since the four cells in the joint distribution must sum to one, the lower bound can be found as the maximum of zero, or one minus the SIV upper bounds on the other three components. In the absence of additional information, these bounds are sharp.

The upper bounds on the marginal distribution given the strong IV assumption equals

(9) $$P[\ Y(1) = 1\ ]\ \#$$

$$\mathrm{Inf}_v\ \{\ P[\ Y(1) = 1 \mid Z = 1, v]\ P[\ Z = 1 \mid v] + P[Y(0) = 1 \mid Z = 0,v]\ P[\ Z = 0 \mid v]\ \}$$
$$+\mathrm{Inf}_v\ \{\ P[\ Y(1) = 1 \mid Z = 1, v]\ P[\ Z = 1 \mid v] + P[Y(0) = 0 \mid Z = 0,v]\ P[\ Z = 0 \mid v]\ \}.$$

The lower bound for P[Y(1)=1] equals one minus the upper bound on P[Y(1)=0]. These bounds may or may not be tighter than the IV bounds in Equation (5).

Balke and Pearl (1997) determine that the sharp bounds on the marginal distribution under the strong IV assumption are the solution to a complex linear programming problem. Combining (5) and (9) we see that these bounds can be expressed using a much simpler formulation. Label the upper and lower IV bounds in (5) as UB_m and $LB_{m,}$ respectively, where the m stands for mean independence IV assumption. Label the upper and lower IV bounds in (9) as UB_s and LB_s respectively, where the s stands for the strong IV assumption. Then, we have

<u>Proposition 1</u>: Given the SIV in Equation 2, the sharp bounds on the marginal distribution of P[y(1) = 1] are

(10) $$\mathrm{Max}\{\ LB_m, LB_s)\ \#\quad P[\ Y(1) = 1\ ]\ \#\quad \mathrm{Min}\{\ UB_m, UB_s)$$

These bounds are identical to the bounds found by Balke and Pearl (1997).

IV.2 Illustration: The Strong IV Assumption's Power

The bounds in Equation 10 reveal that the strong IV assumption can provide additional identifying power over the weak IV assumption. In particular, comparing the upper bounds in (5) and (9) reveals that the strong IV assumptions narrows the upper bound by

(11) $\text{Max}\{ 0,\ \text{Min}_v \{ P[Y(1) = 1 \mid Z = 1, v] P[Z = 1 \mid v] + P[Z = 0 \mid v] \} -$

$[\text{Min}_v \{ P[Y(1) = 1 \mid Z = 1, v] P[Z = 1 \mid v] + P[Y(0) = 1 \mid Z = 0,v] P[Z = 0 \mid v] \}$

$+\text{Min}_v \{ P[Y(1) = 1 \mid Z = 1, v] P[Z = 1 \mid v] + P[Y(0) = 0 \mid Z = 0,v] P[Z = 0 \mid v] \}] \}.$

To illustrate when the strong IV assumption is valuable in practice, it is helpful to assume that V is binary. Then, without loss of generality, there are three interesting cases: all three terms in Equation 11 are minimized when V = 0; the first two terms are minimized when V = 0; the first term is minimized when V = 0. The value of the instrument in these three cases is:

i. <u>Case 1</u>: All three terms in (11) are minimized when V = 0.

(12a) $\text{MAX}\{ 0,\ - P[Y(1) = 1 \mid Z = 1, V=0] P[Z = 1 \mid v = 0] \} = 0$

In this case, the strong IV assumption provides no additional information.

ii. <u>Case 2</u>: The first two terms are minimized when V = 0. The third term is minimized at V = 1.

(12b) $\text{MAX}\{\, 0,\ \ P[Y(0) = 0 \mid Z = 0, V{=}0]\, P[\, Z = 0|V{=}0] \ -$

$\{P[\, Y(1) = 1 \mid Z = 1, V{=}1]\, P[\, Z = 1|\, V{=}1] + P[Y(0) = 0 \mid Z = 0, V{=}1]\, P[\, Z = 0|V{=}1]\,\}$.

=

$\text{MAX}\{\, 0,\ \ P[Y(0) = 0\ ,\ Z = 0\ |V{=}\ 0\ ] - P[\, Y(1) = 1,\ Z = 1|V{=}1] - P[Y(0) = 0,\ Z = 0|\ V{=}1]\ \}$.

In this case, the strong IV assumption may provide additional information. Table 1 below, for example, presents an extreme scenario where the weak IV assumption bounds the marginal distribution of Y(1) whereas the strong IV assumption identifies both the joint and marginal distributions. Balke and Pearl use this same illustration.

iii. <u>Case 3</u>: The first term is minimized when V = 0. The second and third terms are minimized at V = 1.

(12c) $\text{MAX}\{\, 0,\ \ P[\, Y(1) = 1 \mid Z = 1,\ v = 0]\, P[\, Z = 1|\ v = 0]\ + P[\, Z = 0|v = 0]\,\} \ -$

$\{\, 2P[\, Y(1) = 1 \mid Z = 1,\ v = 1]\, P[\, Z = 1|\ v = 1]\ +\ P[\, Z = 0|v = 1]\,\}\ =\ 0$

In this case, the strong IV assumption provides no additional information. To see this, recall that we assume that the upper bound $P[\, Y(1) = 1 \mid Z = 1,\ v\ ]\, P[\, Z = 1|\ v\ ]\ + P[\, Z = 0|v]$ is minimized when v = 0.

IV.3 Multiple Treatments

Not Complete.

V. Extensions to other outcome distributions (is his possible?)

V. Application

Not Complete.

References

Balke, A. and J. Pearl (1997). "Bounds on Treatment Effects from Studies With Imperfect Compliance," Journal of the American Statistical Association, 92, 1171-1177.

Hotz, V.J., C. Mullins and S. Sanders (1997). " Bounding Causal Effects Using Data from Contaminated Natural Experiment: Analyzing the Effects of Teenage Childbearing," Review of Economic Studies, 64, 575-604.

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

Manski, C. (1994), "The Selection Problem," in C. Sims (editor) Advances in Econometrics: Sixth World Congress, Cambridge, UK: Cambridge University Press.

Manski, C. and J. Pepper (2000), "Monotone Instrumental Variables with An Application to the Returns to Schooling," Econometrica,68(4), July 2000, 997-1010

Robins, J. (1989), "The Analysis of Randomized and Non-Randomized AIDS Treatment Trials Using a New Approach to Causal Inference in Longitudinal Studies," in Sechrest, L., H. Freeman, and A. Mulley. eds. Health Service Research Methodology: A Focus on AIDS, NCHSR, U.S. Public Health Service.

Robins, J. and S. Greenland (1996), AComment on Angrist, Imbens, and Rubin=s >Identification of Causal Effects Using Instrumental Variables=,@ Journal of the American Statistical Association, 91, 456-458.

PRELIMINARY DRAFT: PLEASE DO NOT CITE WITHOUT PERMISSION

Table 1

Example 1:

Case ii

P(Y(1)=1,Z=1|V=1)	0.00
P(Y(1)=0,Z=1|V=1)	0.45
P(Y(0)=1,Z=0|V=1)	0.55
P(Y(0)=0,Z=0|V=1)	0.00
P(Y(1)=1,Z=1|V=0)	0.00
P(Y(1)=0,Z=1|V=0)	0.55
P(Y(0)=1,Z=0|V=0)	0.00
P(Y(0)=0,Z=0|V=0)	0.45
P(Z=1|V=1)	0.45
P(Z=0|V=1)	0.55
P(Z=1|V=0)	0.55
P(Z=0|V=0)	0.45

=>

Upper Bounds for P(Y(0),Y(1)|V)

V = 1

		Y(0) 0	Y(0) 1
Y(1)	0	0.45	1.00
	1	0.00	0.55

V = 0

		Y(0) 0	Y(0) 1
Y(1)	0	1.00	0.55
	1	0.45	0.00

=>

Weak IV Bounds for P(Y(0)) and P(Y(1))

0.00	#	P(Y(1)=1)	#	0.45
0.55	#	P(Y(1)=0)	#	1.00
0.55	#	P(Y(0)=1)	#	0.55
0.45	#	P(Y(0)=0)	#	0.45

Strong IV Bounds for P(Y(0)) and P(Y(1))

0.00	#	P(Y(1)=1)	#	0.00
1.00	#	P(Y(1)=0)	#	1.00
0.55	#	P(Y(0)=1)	#	0.55
0.45	#	P(Y(0)=0)	#	0.45

Strong IV Upper Bounds for P(Y(1),Y(0))

		Y(0) 0	Y(0) 1
Y(1)	0	0.45	0.55

1 0.00 0.00