Instrumental Variables

A Study of Implicit Behavioral Assumptions Used In Making Program Evaluations

James Heckman

ABSTRACT

This paper considers the use of instrumental variables to estimate the mean effect of treatment on the treated, the mean effect of treatment on randomly selected persons and the local average treatment effect. It examines what economic questions these parameters address. When responses to treatment vary, the standard argument justifying the use of instrumental variables fails unless person-specific responses to treatment do not influence decisions to participate in the program being evaluated. This requires that individual gains from the program that cannot be predicted from variables in outcome equations do not influence the decision of the persons being studied to participate in the program. In the likely case in which individuals possess and act on private information about gains from the program that cannot be fully predicted by variables in the outcome equation, instrumental variables methods do not estimate economically interesting evaluation parameters. Instrumental variable methods are extremely sensitive to assumptions about how people process information. These arguments are developed for both continuous and discrete treatment variables and several explicit economic models are presented.

You can run from economic models but you can't hide from them.
Derek Neal, 1995

James Heckman is the Henry Schultz Distinguished Service Professor at the University of Chicago and Senior Research Fellow of the American Bar Foundation. *The first draft of this paper was written in August, 1995. This research was supported by grants from the (NSF-SBR-93-24108), the Russell Sage Foundation and the American Bar Foundation. I thank Derek Neal for numerous helpful comments on repeated occasions and Richard Blundell, Lars Hansen, Lance Lochner, Robert Moffitt, Jose Scheinkman and two anonymous referees for their helpful comments. The essential point made in this paper appears in Heckman and Robb (1985). Data used in this article is available beginning November 1997 through October 2000 from the author, Department of Economics, University of Chicago, 1126 East 59th Street, Chicago, IL 60637.*
[Submitted January 1996, accepted January 1997]

I. Introduction

The method of instrumental variables is widely used to evaluate programs and estimate various "treatment effects" including the impact of schooling on earnings. It is routinely invoked when it is suspected that persons sort into programs or schooling levels on the basis of unobserved factors that affect outcomes but are not due to the program or treatment being evaluated. Ability may raise earnings and more able people may go to school but schooling may not raise the earnings of any given person. To resolve this issue, an instrument Z is often sought that determines participation in schooling but that does not directly affect earnings and does not depend on ability. It is easy to check whether Z determines participation. The relationship of Z with unobserved ability is determined by assumption and speculation.

This paper clarifies the implicit behavioral assumptions that underlie application of the method of instrumental variables. Conventional applications of the method assume that the "treatment" being evaluated has the same effect for everyone among persons with a given value of the regressors X. In the simplest case, the assumption is that the effect is the same for all persons. In these cases, the effect of treatment on the treated (Heckman and Robb 1985), the local average treatment effect (Imbens and Angrist 1994) and the effect of treatment on persons selected at random from the population at large are all the same for persons with the same X characteristics. In the more general case when responses to treatment vary among persons with the same X, these three parameters are different, and the method of instrumental variables breaks down unless special assumptions are made about what information the persons or institutions that determine participation act on.

This paper presents the identifying assumptions that justify application of the method of instrumental variables to estimate each of the three conceptually distinct parameters that are frequently confused in the empirical literature. The parameters are defined for both discrete and continuous treatments where appropriate. Simple economic models illustrate the implicit assumptions that are made in applications of the method of instrumental variables. A basic result in this paper is that if responses to treatment vary, and if we are interested in estimating the mean effect of treatment on the treated, or the effect of treatment on randomly selected persons, instrumental variables identify these parameters only when agents do not select into the program on the basis of the idiosyncratic component of their response to the program. This is a strong assumption that forces the analyst to assume either irrationality or ignorance on the part of persons whose behavior is being studied.

I also consider what economic questions these econometric parameters answer. In this regard the local average treatment effect is potentially problematic. The "causal effect" for this parameter is defined by the operation of an instrumental variable external to the outcome equation, and not in terms of parameters of the outcome equation. This is an unusual way to define an economic parameter. Nonetheless, I show that for certain instruments and certain economic environments the local average treatment effect answers a well-posed economic evaluation question.

I first define the three parameters in the standard model of the evaluation problem, which is just the switching regression framework of Quandt (1972, 1988). I start with the more familiar case when treatments are discrete. I consider the case of a continuous treatment variable in a later section.

II. The Evaluation Problem

A person may occupy two potential states, only one of which is realized for any person. Let Y_1 be the outcome in the treated state. Y_0 is the outcome in the untreated state. At any time a person is either in the treated or untreated state but cannot be in both states at the same time. Participation in a program is synonymous with being in the treated state. The gain from going into the program is $\Delta = (Y_1 - Y_0)$.

We cannot form this gain for anyone because one or the other component of the difference is missing. The statistical approach to this problem replaces the missing data on persons using group means or some other group statistics.

Many parameters for evaluating social programs have been proposed. Sometimes, it is of interest to explore the impacts of programs on distributions of outcomes and to determine if it raises the welfare of participants or that of a third party "social planner." This is done in other papers. (See Heckman and Smith 1993; Heckman and Smith 1995; Heckman, Smith and Clements 1997; Heckman and Smith 1997.) Much attention is devoted to the parameter "the mean effect of treatment on the treated." This parameter answers the following question. How does the program change the outcome of participants compared to what they would have experienced if they had not participated? Signify participation by a variable D. For persons who participate, let $D = 1$. For those who do not, let $D = 0$.

The mean change in the outcome attributable to participation in the program for persons with characteristics X is

$$(1)\quad E(\Delta|D = 1, X) = E(Y_1 - Y_0|D = 1, X).$$

We know or can reliably estimate $E(Y_1|D = 1, X)$. This is what participants experience. We don't know $E(Y_0|D = 1, X)$, what participants would have experienced had they not participated.

A second counterfactual that many confuse with the mean effect of "treatment on the treated" is the effect of "randomly assigning a person in the population to the program." That counterfactual is

$$(2)\quad E(\Delta|X) = E(Y_1 - Y_0|X).$$

Neither component of this mean has a sample analogue unless there is universal participation or nonparticipation in the program, or participation is randomly determined and there is full compliance with the randomization regime.

The intuitively appealing counterfactual (2) is very difficult to estimate. Picking a millionaire at random to participate in a training program for low skilled workers, or making an idiot into a PhD may be intriguing thought experiments but are usually neither policy relevant nor feasible. They are not policy relevant because interest centers on the effects of programs on intended recipients—not on persons for whom the program was never intended. It is not a feasible random-assignment strategy because millionaires would never agree to participate in such a training program even if they were offered the chance to do so, and few idiots would be able to attain the PhD in most fields.[1]

A third counterfactual is the effect of treatment on persons at the margin of being treated. This is a local version of treatment on the treated and requires that the relevant margin be specified. It is formally defined in Section V.

Understanding the differences among these counterfactuals is of central importance in understanding competing approaches used in the evaluation literature. Previous practice in the econometric evaluation community (for example, Ashenfelter 1978, LaLonde 1986, Ashenfelter and Card 1985) assumes a special model in which all three counterfactuals are the same.

III. Constructing Counterfactuals

While the linear regression model is commonly used, it is actually simpler and at the same time more general to take a nonparametric approach and condition on regressors X. Define

$$E(Y_1|X) = \mu_1(X)$$
$$E(Y_0|X) = \mu_0(X).$$

Thus we may write

3(a) $Y_0 = \mu_0(X) + U_0$

3(b) $Y_1 = \mu_1(X) + U_1$

where $E(U_0|X) = 0$ and $E(U_1|X) = 0$. In the familiar regression setting, $\mu_0(X) = X\beta_0$ and $\mu_1(X) = X\beta_1$, but our results apply more generally.

Observed outcome Y can be written as

$$Y = DY_1 + (1 - D)Y_0$$

so Y is either Y_1 or Y_0. If we insert (3a) and (3b) into this expression, we obtain

(4) $Y = \mu_0(X) + D(\mu_1(X) - \mu_0(X) + U_1 - U_0) + U_0.$

This is a "two regime" or "switching regression" model (see Quandt 1972, 1988). Labor economists sometimes call it a Roy model (see, for example, Heckman and Sedlacek 1985 or Heckman and Honoré 1990).

The term multiplying D is the gain from the program. The gain has two components: $\mu_1(X) - \mu_0(X)$ is the gain for the average person with characteristics X in the population, and $U_1 - U_0$ is the idiosyncratic gain for a particular person. Note that these unobservables may be observed by the person or persons deciding

1. Because of agent self-selection, random assignment of eligibility only identifies the effect of treatment on the treated unless further assumptions, beyond the validity of random assignment, are assumed. See Heckman (1996).

participation in the program. They are unobserved by the social scientist trying to estimate the impact of the program.

In this notation, the average gain for persons with characteristics X is

$$E(\Delta|X) = \mu_1(X) - \mu_0(X).$$

This is the effect of placing an average person with characteristics X in the population at large into the program.

The effect of treatment on the treated for persons with characteristics X is

$$E(\Delta|X,D = 1) = \mu_1(X) - \mu_0(X) + E(U_1 - U_0|X,D = 1).$$

This expression differs from the previous one by the additional term $E(U_1 - U_0|X, D = 1)$. $E(\Delta|X) = E(\Delta|X, D = 1)$ when $E(U_1 - U_0|X, D = 1) = 0$. This can happen if $U_1 - U_0 = 0$ or if agents either do not know $U_1 - U_0$ or do not act on it. The term added to the difference in population means $(\mu_1(X) - \mu_0(X))$ is the gain of participants over the average gain that would be experienced by the entire population with characteristics X. I define the local average treatment effect in Section V.

We may rewrite Equation 4 in terms of these two parameters:

$$(5) \quad Y = \mu_0(X) + D[E(\Delta|X)] + \{U_0 + D(U_1 - U_0)\}$$

and

$$(6) \quad Y = \mu_0(X) + E(\Delta|X,D = 1) + \{U_0 + D[U_1 - U_0 - E(U_1 - U_0|X,D = 1)]\}.$$

From Equation 5, in a regression of Y on D, the coefficient on D is

$$E(Y|X,\mathrm{D} = 1) - E(Y|X,\mathrm{D} = 0)$$
$$= E(\Delta|X) + E(U_1 - U_0|X,D = 1) + [E(U_0|X,D = 1) - E(U_0|X,D = 0)]$$
$$= E(\Delta|X,D = 1) + E(U_0|X,D = 1) - E(U_0|X,D = 0].$$

The coefficient differs from $E(\Delta|X)$ by the amount

$$E(U_1 - U_0|X,D = 1) + E(U_0|X,D = 1) - E(U_0|X,D = 0).$$

It differs from $E(\Delta|X, D = 1)$ by

$$E(U_0|X,D = 1) - E(U_0|X,D = 0).$$

This term is the mean selection bias. This bias tells us how the outcome in the base state differs between program participants and nonparticipants. Absent any general equilibrium effects of the program on nonparticipants, such differences cannot be attributed to the program.

The parameters $E(\Delta|X, D = 1)$ and $E(\Delta|X)$ coincide when the mean change in the unobservable conditional on D is zero, namely, $E(U_1 - U_0|X, D = 1) = 0$. We now examine in detail the two special cases when this condition is satisfied and $E(\Delta|X) = E(\Delta|X, D = 1)$.

In the first case, there are no unobservable components of the gain. This model—called the "dummy endogenous variable model" (see Heckman 1978)—is widely used in applied work (see Ashenfelter 1978; Ashenfelter and Card 1985; LaLonde 1986). It assumes that conditional on X, the effect of program participation is the same for everyone. This is sometimes called the common coefficient model.

The second case where $U_1 \neq U_0$ is more subtle. In this case $U_1 - U_0$ or information correlated with or dependent on it does not determine who goes into the program. Suppose, for example, that at the time people go into the program they do not know $U_1 - U_0$. Their best forecast of $U_1 - U_0$ may be zero. Then if their expectation of $U_1 - U_0$ is typical of that of the entire population, $E(U_1 - U_0|X, D = 1) = 0$, and $E(\Delta|X) = E(\Delta|X, D = 1)$. This case is analogous to the "random coefficients" model of traditional econometrics.[2] Ex ante, persons with the same X have identical expectations of gain. Ex post, people respond differently to training. Observe that in either case, the problem of estimating $E(\Delta|X)$ or $E(\Delta|X, D = 1)$ using the difference in outcomes between participants and nonparticipants, is the standard econometric endogeneity problem that D is stochastically dependent on U_0.

Irrespective of whether $E(U_1 - U_0|X, D = 1) = 0$ in Equation 6, the component of the error term interacted with D has mean zero. That is, $E(D(U_1 - U_0) - E(U_1 - U_0|X, D = 1)|X, D) = 0$, because when $D = 1$,

$$E(U_1 - U_0 - E(U_1 - U_0|X, D = 1)|X, D = 1)$$
$$= E(U_1 - U_0|X, D = 1) - E(U_1 - U_0|X, D = 1) = 0.$$

Therefore D is uncorrelated with $D(U_1 - U_0 - E(U_1 - U_0|X, D = 1))$. Even if $E(\Delta|X) \neq E(\Delta|X, D = 1)$, for estimating $E(\Delta|X, D = 1)$ the problem is the standard one of correlation between D and U_0. However for estimating $E(\Delta|X)$, an additional source of bias arises from the dependence between D and $U_1 - U_0$ (Heckman and Robb 1985).

IV. The Method Of Instrumental Variables

A standard method for estimating parameters in econometrics is the method of instrumental variables. Instrumental variables must satisfy two basic conditions. They must be mean-independent of the error terms of Equations 5 and 6 that is,

$$(C-1-a)\ E[U_0 + D(U_1 - U_0)|X, Z] = 0 \text{ (an identifying condition for } E(\Delta|X))$$

or

$$(C-1-b)\ E[U_0 + D(U_1 - U_0 - E((U_1 - U_0)|D = 1, X))|X, Z] = 0$$
$$\text{(an identifying condition for } E(\Delta|X, D = 1)).$$

These statistical assumptions are not innocuous and it will be demonstrated that they rest on implicit behavioral assumptions that are much stronger than what is

2. In the case considered in this paper, however, regressor D is correlated with U_0. A testable restriction of this model when U_0 is mean independent of D and X, discussed in Heckman and Robb (1985) and Heckman, Smith, and Clements (1997) is that $\text{Var}(Y_1|D = 1, X) > \text{Var}(Y_0|D = 0, X)$. Heckman, Smith, and Clements (1997) test and accept this restriction for a job training program. They also discuss modification of this test for the case where $E(U_0|D, X) \neq 0$.

required when the response to treatment is homogeneous among persons with the same value of X.

A second required condition is that D depends on Z:

$$(C-2)\ E(D|X,Z) = \Pr(D = 1|X,Z)$$

where the probability is a nontrivial function of Z given X. This requires that there be independent variation in Z conditioning on X. In particular, it is required that for each X, there be two values of Z, $z \neq z'$, which produce different values for the probability. This assumption can be checked by estimating the probability to see if it depends on Z. Assumptions $(C-1-a)$ and $(C-1-b)$ are not testable when there is only one instrumental variable.

As a consequence of these conditions, the dependence of Y on Z operates only through D. This excludes dependence of the parameters of interest on Z. Thus for $E(\Delta|X)$, $(C-1-a)$ and $(C-2)$ imply

$$E(Y|X,Z) = \mu_0(X) + E(\Delta|X)\Pr(D = 1|X,Z).$$

For $E(\Delta|X, D = 1)$, $(C-1-b)$ and $(C-2)$ imply

$$E(Y|X,Z) = \mu_0(X) + E(\Delta|X,D = 1)\Pr(D = 1|X,Z).$$

Instruments are variables that "don't belong in the population outcome equation" but that "belong" in the equation predicting program participation.

The population instrumental variable equation for $E(\Delta|X)$ is, under $(C-1-a)$ and $(C-2)$,

$$(7)\quad E(\Delta|X) = \frac{E(Y|X,Z = z) - E(Y|X,Z = z')}{\Pr(D = 1|X,Z = z) - \Pr(D = 1|X,Z = z')}$$
$$= \frac{E(\Delta|X)[(\Pr(D = 1|X,Z = z) - \Pr(D = 1|X,Z = z')]}{[\Pr(D = 1|X,Z = z) - \Pr(D = 1|X,Z = z')]}$$

where $z \neq z'$ and the denominator is not zero. Replacing population means with sample means produces the instrumental variable estimator which, under standard conditions, converges to $E(\Delta|X)$. The parameter $E(\Delta|X)$ is not a function of Z, so the right hand side is not either.

The population instrumental variable equation for $E(\Delta|X, D = 1)$ is, under $(C-1-b)$ and $(C-2)$,

$$(8)\quad E(\Delta|X,D = 1) = \frac{E(Y|X,Z = z) - E(Y|X,Z = z')}{\Pr(D = 1|X,Z = z) - \Pr(D = 1|X,Z = z')}.$$

Again, the right hand side ratio does not depend on Z because of assumption $(C-1-b)$.[3,4]

3. A chain rule interpretation of IV notes that if $E(Y|X, Z)$ and $\Pr(D = 1|X, Z)$ are differentiable in Z the IV estimator is just the ratio of the derivatives

$$\lim_{z\to z'} \text{IV} = \frac{\partial E(Y|X,Z = z')}{\partial Z} \Big/ \frac{\partial \Pr(D = 1|X,Z = z')}{\partial Z}$$

which does *not* depend on Z under the assumptions that justify application of the estimator to identify $E(\Delta|X, D = 1)$ or $E(\Delta|X)$. The parameter is the ratio of the change in the conditional expectations with respect to Z to the ratio of the change of the probability with respect to Z.

In a model where $U_1 - U_0$ is not zero and is not a determinant of D, in other words, where $(U_1 - U_0)$ is statistically independent of X, Z, D, and hence

$$\Pr(D = 1|X, Z, U_1 - U_0) = \Pr(D = 1|X, Z, Y_1 - Y_0) = \Pr(D = 1|X, Z),$$

we can also ignore the component $D(U_1 - U_0)$ in (4) in forming instrumental variable equations because

$$E(D(U_1 - U_0)|X, Z) = E((U_1 - U_0)|X, Z, D = 1)\Pr(D = 1|X, Z) = 0$$

because $E(U_1 - U_0|X, Z, D = 1) = 0$. All we really require is mean independence: $E(U_1 - U_0|X, Z, D = 1) = E(U_1 - U_0) = 0$. Thus in the two cases where $E(\Delta|X) = E(\Delta|X, D = 1)$, where $U_1 - U_0 \equiv 0$ or $U_1 - U_0$ cannot be forecast by Z or D, we can use conventional textbook instrumental variable methods to identify the parameter $E(\Delta|X) = E(\Delta|X, D = 1)$.[5]

What about more general cases? Consider Equation 6 with associated treatment parameter $E(\Delta|X, D = 1)$. Because the only source of dependence between the error term and D is through U_0 and not $D(U_1 - U_0 - E(U_1 - U_0|X, D = 1))$, the instrumental variable method looks promising. If assumptions $(C-1-b)$ and $(C-2)$ are satisfied, the IV moment conditions can be used to identify this parameter.

In general such instrumental variables are difficult to find. If the unobservable $U_1 - U_0$ determines participation so $\Pr(D = 1|X, Z, U_1 - U_0) \neq \Pr(D = 1|X, Z)$ then by Bayes' rule $E(U_1 - U_0|X, Z, D = 1) \neq 0$, so from Equation 5,

$$E(\Delta|X, Z, D = 1) = E(\Delta|X) + E(U_1 - U_0|X, Z, D = 1)$$

functionally depends on Z.[6] If individuals select into the program on the basis of the unobservables in the outcome equation or on the basis of the variables that are (stochastically) dependent on the gain in unobservables, condition $(C-1-b)$ will not be satisfied. (See Heckman and Robb 1985, 1986.) Z determines the

4. Observe that (7) is valid even if (C-1-a) is weakened to $E[U_0 + D(U_1 - U_0)\ |X, Z] = M(X)$ a function of X and (8) is valid even if (C-1-b) is weakened to $E[U_0 + D(U_1 - U_0 - E(U_1 - U_0|D = 1, X))|X, Z] = K(X)$. These functions difference out for each X, and need not be zero.

5. These methods are also justified in a third case where $U_1 - U_0$ cannot be perfectly forecast by Z but can be forecast by X.

6. From Bayes' rule,

$$\Pr(D = 1|X, Z, U_1 - U_0)f(U_1 - U_0|X, Z)f(X, Z) = f(U_1 - U_0|X, Z, D = 1)\Pr(D = 1|X, Z)f(X, Z)$$

where $f(U_1 - U_0|X, Z)$ is the conditional density of $(U_1 - U_0)$ given X, Z, $f(U_1 - U_0|X, Z, D = 1)$ is the conditional density of $U_1 - U_0$ given $X, Z, D = 1$ and $f(X, Z)$ is the joint density of (X, Z). Then for all values of (X, Z) so that $\Pr(D = 1|X, Z) \neq 0$, and $f(X, Z) > 0$,

$$f(U_1 - U_0\,|X, Z, D = 1) = \frac{\Pr(D = 1|\,X, Z, U_1 - U_0)f(U_1 - U_0\,|X, Z)}{\Pr(D = 1|X, Z)}$$

so the conditional mean of $U_1 - U_0$, $E(U_1 - U_0|X, Z, D = 1)$ which is computed with respect to this density, is in general a function of Z. Observe that this dependence exists even if $U_1 - U_0$ is independent of X, Z so $f(U_1 - U_0|X, Z) = f(U_1 - U_0)$. If $U_1 - U_0$ is independent of X, Z and $U_1 - U_0$ does not determine $\Pr(D = 1|X, Z, U_1 - U_0)$ so $\Pr(D = 1|X, Z, U_1 - U_0) = \Pr(D = 1|X, Z)$, then $U_1 - U_0$ is independent of X, Z, D.

parameter and is not a valid instrument for $E(\Delta|X, D = 1)$.[7] If condition $(C-1-b)$ is not satisfied, neither will condition $(C-1-a)$ be satisfied, so $E(\Delta|X)$ will not be identified. Thus, even if $E(U_0|X, Z) = 0$, the requirement that $E(D(U_1 - U_0)|X, Z) = E(U_1 - U_0|X, Z, D = 1)\Pr(D = 1|X, Z) = 0$ will not in general be satisfied even if $E(U_1 - U_0|X, Z, D = 1) = E(U_1 - U_0|X, D = 1)$.

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

Consider an example that is often cited as a triumph for the application of the method instrumental variables. Draft lottery numbers are sometimes alleged to be ideal instrumental variables for identifying the effect of military service on earnings (Angrist 1990). The 1969 U.S. lottery randomly assigned different priority numbers to persons with different birth dates. The higher the number in the draft, the less likely was a person to be drafted. Persons with high numbers were virtually certain to be able to escape the draft. Letting 1 denote military service and 0 civilian service, if persons partly anticipate gain in $U_1 - U_0$, or base their decisions to go into the military on variables correlated with unobservable components $U_1 - U_0$, persons with high Z for whom $D = 1$ (they serve in the military) are likely to have high values of $U_1 - U_0$. This violates assumption

7. If $(C-1-b)$ is violated, then

$$E(Y|Z = z) = \mu_0(X) + E(\Delta|X, Z = z, D = 1)\Pr(D = 1 X, Z = z)$$
$$E(Y|Z = z') = \mu_0(X) + E(\Delta|X, Z = z', D = 1)\Pr(D = 1 X, Z = z')$$

so

$$E(Y|Z = z) - E(Y|Z = z') = E(\Delta|X, Z = z, D = 1)\Pr(D = 1|X, Z = z) - E(\Delta|X, Z = z', D = 1)\Pr(D = 1 X, Z = z')$$

so

$$\frac{E(Y|Z = z) - E(Y|Z = z')}{\Pr(D = 1|X, Z = z) - \Pr(D = 1|X, Z = z')} = \frac{E(\Delta|X, Z = z, D = 1)\Pr(D = 1|X, Z = z) - E(\Delta|X, Z = z', D = 1)\Pr(D = 1 X, Z = z')}{\Pr(D = 1|X, Z = z) - \Pr(D = 1|X, Z = z')}$$

a weighted average of the treatment effects. In the limit as $z \to z'$ if the functions of Z are differentiable, letting "IV" be the IV moment condition,

$$\lim_{z \to z'} \text{IV} = \frac{\dfrac{\partial E(Y|Z = z')}{\partial Z}}{\dfrac{\partial \Pr(D = 1|\, Z = z')}{\partial Z}} = E(\Delta|X, Z = z', D = 1) + \Pr(D = 1|X, Z = z')\frac{\dfrac{\partial E(\Delta|X, Z = z', D = 1)}{\partial Z}}{\dfrac{\partial \Pr(D = 1|X, Z = z')}{\partial Z}}.$$

The final term is the probability-weighted effect of a change in Z on the mean effect of treatment on the treated.

$(C-1-b)$ because $E(U_1 - U_0|D = 1, X, Z)$ will depend on Z, and makes the birth date number an invalid instrument for identifying $E(\Delta|X, D = 1)$. It is plausible that the persons who are deciding to go into the military know more about their gains from military service than analysts using standard data sets. If this information is at all useful in predicting their gain from going into the military, the draft number is not a valid instrument. Persons who have high Z who go into the military are more likely to have high $U_1 - U_0$. That is, they at least partly anticipate substantial gains to entering the military. Observe that if $(C-1-b)$ fails, $(C-1-a)$ must fail as well because then $E(D(U_1 - U_0)|X, Z) = E(U_1 - U_0|D = 1, X, Z) \Pr(D = 1|X, Z) \neq 0$ and hence the instrument also fails to identify $E(\Delta|X)$.[8,9]

As a second example, it is sometimes suggested that cross-state variation in welfare benefits can be used as instrumental variables for estimating the effect of "treatment on the treated" for participants in training programs. Suppose that Y_0 refers to the earnings of untrained low-skill persons. Y_1 is their earnings if trained. Parameters of welfare benefit functions, Z, do not plausibly enter $\mu_0(X)$ or $\mu_1(X)$. But they could enter $E(U_1 - U_0|X, Z, D = 1)$. If more generous welfare schemes discourage participation in the training program because they induce people to stay out of the market, then higher values of $U_1 - U_0$ would tend to be found among program participants in high benefit states if program

8. For another reason, the draft lottery number is a poor instrument for identifying $E(\Delta|X)$ or $E(\Delta|X, D = 1)$. Switching from a regime of a capricious draft to a lottery reduces uncertainty and is likely to change the investment behavior of persons at all levels of Z. In this instance, the switch from a draft to a lottery affects both $E(\Delta|X, D = 1)$ and $E(\Delta|X)$ because it fundamentally alters schooling and job training investment decisions and their payoffs. Thus, knowing how military service affects earnings during the period of a lottery would not be informative about how military service affected earnings during the period of an ordinary draft. There is yet another reason why the draft lottery is a poor instrument. Z is likely to be an X. Persons with high Z (a low chance of being drafted) are likely to be more attractive to employers investing in their workers. A person unlikely to be drafted is likely to be a better investment because he is less likely to be removed from the firm to perform military service. This causes $(C-1-b)$ to be violated because Z is really an X.

9. It might be thought that because

$$E[U_1 - U_0 - E(U_1 - U_0|X, D = 1)|X, D = 1] = 0,$$

it follows that $E[U_1 - U_0 - E(U_1 - U_0|X, D = 1)|X, Z, D = 1] = 0$. This is not true. In general

$$E(U_1 - U_0|X, Z, D = 1) \neq E(U_1 - U_0|X, D = 1).$$

Conditioning more finely on X and Z does not produce the same result as conditioning on X. Then even if

$$E(U_0|X, Z) = 0,$$

it does not follow that $(C-1-b)$ is satisfied. In this case, Z is not a valid instrument for identifying $E(\Delta|X, D = 1)$. For similar reasons, $(C-1-a)$ is unlikely to be satisfied even if

$$E(U_0|X, Z) = 0$$

because

$$E[D(U_1 - U_0|X, Z)] = E[U_1 - U_0|X, Z, D = 1] \Pr(D = 1|X, Z)$$

does not equal zero. In this case Z is not a valid instrument for identifying $E(\Delta|X)$ either.

participants enter the program with at least partial knowledge of $U_1 - U_0$. Then assumption $(C-1-b)$ is violated, and cross-state variation in benefits do not identify $E(\Delta|X, D = 1)$ or $E(\Delta|X)$ through the method of instrumental variables.

A third example is the problem of evaluating the impact of unionism on wage rates. Robinson (1989) uses a test developed by Heckman and Robb (1985, p. 196) and determines that in his sample decisions to join unions are not made on the basis of unobservable wage gains. Factors other than the unobservables predicting wage gains determine union membership. His evidence does not rule out the possibility that union membership is made on the basis of $\mu_1(X) - \mu_0(X)$. It does rule out the possibility that the unobservables in wage equations determine union membership. For this problem the method of instrumental variables correctly identifies the effect of unionism on wages among those who choose to enter unions.

V. The "Local Average Treatment Effect"

Imbens and Angrist (1994) introduce a new parameter into the evaluation literature: the effect of treatment on those who change state in response to a change in Z. More precisely, their parameter is $E(Y_1 - Y_0|D(z) = 1, D(z') = 0)$ where $D(z)$ is the conditional random variable D given $Z = z$, and where z' is distinct from z, so $z \neq z'$. This parameter is termed *LATE* for the Local Average Treatment Effect. The parameter has several nonstandard features. It is *defined* by variation in an instrumental variable that is external to the outcome equation. Unlike the instrumental variables discussed in the preceding section, in *LATE*, different instruments *define* different parameters. When the instruments are indicator variables that denote different policy regimes, *LATE* has a natural interpretation as the response to policy changes for those who change participation status in response to the change. When the instruments refer to personal or neighborhood characteristics used to predict an endogenous variable, say schooling in an earnings equation, *LATE* often has a less clear cut interpretation. If distance to the nearest school is the instrument, *LATE* estimates the effect of variation in distance on the earnings gain of persons who are induced to change their schooling status as a consequence of commuting costs that vary within a specified range. If a personal characteristic is used as an instrument (for example, family income), the parameter defines the marginal change in the outcome with respect to the sample variation in family income among those who would have changed their state in response to the sample variation in family income.

There is another nonstandard feature of *LATE*. For any given instrument, *LATE* is defined on an unidentified hypothetical population: persons who would certainly change from 0 to 1 if Z is changed. For different values of Z and for different instruments, the *LATE* "parameter" changes and the population for which it is defined changes.

To define the *LATE* parameter more precisely, let $D(z)$ be the conditional random variable D given $Z = z$. (Conditioning on X is kept implicit in this section). Because $D(z)$ is defined conditional on a particular realization of $Z = z$, it is independent of Z.[10] Imbens and Angrist (1994) assume

IA-1 (Y_0, Y_1, $D(z)$) are independent of Z and $\Pr(D = 1|Z = z)$ is a nontrivial function of Z. (These random variables are understood to be defined conditional on X).

As a consequence of this assumption, for a given person (with fixed Y_1, Y_0), recalling that for $Z = z$, $Y = Y_0(1 - D(z)) + Y_1 D(z)$, so

$$(9)\quad E(Y|Z = z) - E(Y|Z = z')$$
$$= E[D(z)Y_1 + (1 - D(z))Y_0|Z = z] - E[D(z')Y_1 + (1 - D(z'))Y_0|Z = z']$$
$$= E((D(z) - D(z'))(Y_1 - Y_0)).$$

The final step follows from assumption IA-1 and depends crucially on the conditional independence of Y_1, Y_0 and $D(z)$ from Z.

In the Imbens-Angrist thought experiment, all of the random variables in the expression are defined for the same person. Thus for different values of $Z = z$, Y_1 and Y_0 do not change and $\{D(z)\}_{z \in \text{support of} Z}$ is a collection of not necessarily independent random variables produced by changing z and not by changing any other random variable or by changing them in a way specified in assumption IA-2, below. In terms of the index model of discrete choice theory where index IN can be written in terms of the index function $g(Z, V)$, which may be a net profit or net utility function IN $= g(z, V)$

$$(10)\quad D = 1(g(z, V) \geq 0)$$

and V is a random variable. V stays fixed in the Imbens-Angrist thought experiment while z is varied.

From Equation 9 it follows that

$$(11)\quad E(Y|Z = z) - E(Y|Z = z')$$
$$= E(Y_1 - Y_0|D(z) - D(z') = 1)\Pr(D(z) - D(z') = 1)$$
$$+ E(Y_1 - Y_0|D(z) - D(z') = -1)\Pr(D(z) - D(z') = -1).$$

Imbens and Angrist call $E(Y_1 - Y_0|D(z) - D(z') = 1)$ and $E(Y_1 - Y_0|D(z) - D(z') = -1)$ "causal" parameters. In general, these parameters depend on the particular choice of the z and z' as well as X. Factors external to the outcome equation define the *LATE* parameter and a different parameter is produced for each choice of z and z'. If there are multiple instruments, there are multiple parameters. Additional instruments do not improve efficiency as they would in the models considered in previous sections of this paper and in standard "policy invariant" structural models. They instead define different parameters.

Imbens and Angrist are not interested in parameter (11) although it answers the question of how changes in Z change overall outcomes. To identify one of their "causal" parameters, they invoke a second assumption about their hypothetical random variables:

10. For two random variables (J, K) let f be the density (or frequency). Then $f(J, K) = f(J|K)f(K)$ so J given K is statistically independent of K although $f(J|K)$ may be functionally dependent on K.

IA-2 For all z, z' in the support of z, either $D(z) \geq D(z')$ for all persons or $D(z) \leq D(z')$ for all persons.

The variation across z and z' is made holding the error term constant. This condition makes either $\Pr(D(z) - D(z') = 1)$ or $\Pr(D(z) - D(z') = -1)$ zero for everyone. Thus, the effect of a change in Z is to shift people in one sector or the other but not both. Suppose $D(z) \geq D(z')$, then $\Pr(D(z) - D(z') = -1) = 0$ and using (11) we obtain

$$(12) \quad E(Y_1 - Y_0 | D(z) - D(z') = 1) = \frac{E(Y|Z = z) - E(Y|Z = z')}{\Pr(D = 1|Z = z) - \Pr(D = 1|Z = z')}.$$

Because this parameter is *defined* in terms of population moments, it can be consistently estimated by instrumental variables methods replacing population moments by sample moments.

Comparing (12) with (7) or (8) reveals that *LATE* looks like what the standard IV converges to except for one important difference. The *LATE* parameter is now z dependent. Both *LATE* and the parameters (7) or (8) are identified by taking the ratio of the change in the outcome induced by Z and dividing by the change in the probability of being in sector 1 induced by $Z = z$. Parameters (7) or (8) do not depend on Z while the *LATE* parameter does.

Observe further that if conditions $(C-1-a)$ and $(C-2)$ are satisfied, *LATE* identifies $E(\Delta|X)$. If $(C-1-b)$ and $(C-2)$ are satisfied the population moment condition used to define *LATE* identifies $E(\Delta|X, D = 1)$.

Condition IA-2 is satisfied if (10) characterizes choices. It is also satisfied if

$$\text{IN} = g(z, V_z) \text{ and } D(z) = 1(\text{IN} > 0|Z = z)$$

characterizes participation in the program being evaluated provided that g is increasing in z, V_z is increasing in z and g is increasing in V_z. This would be satisfied in the case of a scalar z if $z > z'$, and $\sigma(z)$ is a random variable if

$$V_z = V_{z'} + \sigma(z)$$

where $\sigma(z) > 0$, for $z > z'$. If, however, $\sigma(z)$ is permitted to be both positive and negative, condition IA-2 would not be satisfied.[11]

11. The Roy model estimated by Heckman and Sedlacek (1985) has a decision rule of the form of Equation 10

$$\text{IN} = Y_1 - Y_0 + k(z),$$

and if $k(z)$ is monotonic in z produces a model consistent with IA-2. Assume that Z is independent of $(Y_1 - Y_0)$ so that the conditions of IA-1 are satisfied. $Y_1 - Y_0 = V$ is fixed and different realizations of Z are considered in the Imbens-Angrist thought experiment that defines their parameter.

In this set up, the event $D(z) - D(z') = 1$ is described by the inequalities

$$Y_1 - Y_0 + k(z) > 0 \text{ and } Y_1 - Y_0 + k(z') < 0$$

(assume that $Y_1 - Y_0$ are continuous random variables) so

$$-k(z') > Y_1 - Y_0 > -k(z)$$

and the model induces a partition of $Y_1 - Y_0$. Now the *LATE* "causal parameter" is

$$E(Y_1 - Y_0|D(z) - D(z') = 1) = E(Y_1 - Y_0| -k(z) < Y_1 - Y_0 < -k(z'))$$

When parameters are defined to be instrument dependent, and they are defined for unobserved subsets of the population (those who would have changed state if their Z were changed while their unobservables were held fixed), it is no longer clear what interesting policy question they answer. However, it is clear what econometric problem this redefinition of the parameter of interest avoids. As noted by Heckman (1990), it is necessary to have subsets of the support of Z where there is no selection bias to identify $E(\Delta | X, D = 1)$. By redefining the parameter of interest, Imbens and Angrist avoid this problem, but at the cost of defining an instrument-dependent parameter that will vary across samples and across instruments even within a sample. We now turn to the question of what economic question is addressed by LATE and the other econometric evaluation parameters.

VI. What Economic Policy Questions Do The Parameters Answer?

If assumptions IA-1 and IA-2 are ignored, a parameter similar to *LATE* answers an interesting policy question. Consider a program with a voluntary component. Choice of sector 1 is tantamount to participation in the voluntary component of the program. Choice of sector 0 is tantamount to nonparticipation. These terms are used generally and include failure to comply with a mandatory program, that is, voluntary nonparticipation. If $Z = z$ is the level of a policy, and it is changed, then if Y is income, and Z affects both participation choices and average incomes of participants and nonparticipants,

$$E(Y|Z = z) - E(Y|Z = z')$$

is the change in per capita income resulting from the change in the policy from z' to z.[12] In terms of aggregate income accounting, this parameter measures the per capita change in income from the policy. There is no reason to assume that policies shift people unidirectionally to participate or not to participate nor is there any reason to expect that outcomes for participants and nonparticipants do not depend on Z. Thus neither IA-1 nor IA-2 are especially compelling assumptions in the context of evaluating the overall impact of a policy.

By placing some additional structure on the problem, we can link *LATE* to a criterion that is widely used in the literature on microeconomic program evaluation and also establish a link with the discrete choice literature. Assume a binary choice random utility framework. Suppose that there are two choices (participation or nonparticipation) and that agents make their program participation choices based on net utility. Assume that policies affect participant utility through an additively-separable term $k(Z)$ that is assumed scalar and differentiable. Net welfare W to the agent is

$$W = X + k(Z)$$

which clearly depends on the choice of z and z'. This example illustrates the point that statistical independence of two random variables does not imply their functional independence.

12. The argument in this section is developed in much greater detail in Heckman and Smith (1997).

where k is monotonic in Z and where the joint distributions of (Y_1, X) and (Y_0, X) are $F(y_1, x)$ and $F(y_0, x)$, respectively. In the special case of the original Roy model, $W = Y_1 - Y_0$. For simplicity, Y_1, Y_0, and X are assumed to be continuous random variables. In the general case,

$$D = 1(W \geq 0 | Z = z) = 1(X \geq -k(z) | Z = z).$$

The proportion of people choosing to participate ($D = 1$) is

$$\Pr(D = 1 | Z = z) = \Pr(W \geq 0 | Z = z) = \int_{-k(z)}^{\infty} f(x) dx.$$

The proportion of people choosing not to participate ($D = 0$) is

$$\Pr(D = 0 | Z = z) = \Pr(W < 0 | Z = z) = \int_{-\infty}^{-k(z)} f(x) dx.$$

Total output per capita is

$$E(Y_1 | D = 1, Z = z) \Pr(D = 1 | Z = z) + E(Y_0 | D = 0, Z = z) \Pr(D = 0 | Z = z) - c(z)$$
$$= \int_{-\infty}^{\infty} y_1 \int_{-k(z)}^{\infty} f(y_1, x | z) dx dy_1 + \int_{-\infty}^{\infty} y_0 \int_{-\infty}^{-k(z)} f(y_0, x | z) dx dy_0 - c(z),$$

where $c(z)$ is the direct cost of the policy. This is just the sum of the mean outputs in each sector multiplied by the proportion of people in each sector less the direct cost $c(z)$. Under standard conditions, we may differentiate under the integral sign to obtain the following expression for the marginal change in output with respect to a change in (z) (see Royden 1968).

$$\Delta(z)$$
$$= k'(z) f_x(-k(z)) \{E(Y_1 | D = 1, X = -k(z), Z = z) - E(Y_0 | D = 0, X = -k(z), Z = z)\}$$
$$+ \left[\int_{-\infty}^{\infty} y_1 \int_{-k(z)}^{\infty} \frac{\partial f(y_1, x | z)}{\partial z} dx dy_1 + \int_{-\infty}^{\infty} y_0 \int_{-\infty}^{-k(z)} \frac{\partial f(y_0, x | z)}{\partial z} dx dy_0 \right] - c'(z).$$

where $f_x(-k(z))$ is the density of X evaluated at $X = -k(z)$. This model has a well-defined margin: $X = -k(z)$. The first set of terms corresponds to the gain arising from the movement of persons at the margin weighted by the proportion of the population at the margin, $f_x(-k(z))$. This term is the net gain from switching from nonparticipant to participant status. The term in braces is a limit form of the "local average treatment effect." The second set of terms is the change in output within the two sectors resulting from the policy change. This term is ignored in many evaluation studies and is not incorporated in the definition of the "causal parameters" defined by Imbens and Angrist. It describes how people who do not switch sectors are affected by the policy. The third term is the direct marginal social cost of the policy change, which is rarely estimated. At an optimum, $\Delta(z) = 0$, provided standard second order conditions are satisfied. Marginal benefit should equal marginal cost.

Observe that the local average treatment effect is simply the effect of treatment on the treated for persons at the margin ($X = -k(z)$)

$$E(Y_1|D = 1, X = -k(z), Z = z) - E(Y_0|\, D = 0, X = -k(z), Z = z)$$
$$= E(Y_1 - Y_0\,|D = 1, X = -k(z), Z = z).$$

The proof of this result is immediate once it is recognized that the set $X = -k(z)$ is the indifference set. Thus, the *LATE* parameter is a marginal version of the conventional "treatment on the treated" evaluation parameter for gross outcomes except now the parameter depends on Z. This parameter is only one of the ingredients required to produce an evaluation of social welfare if per capita GNP (Gross National Product) is taken as the appropriate measure of social welfare. Alternatively, *LATE* estimates only the gain from the policy to the persons induced to change participation status by the policy, and ignores the effect of the variation in policy parameter z on inframarginal persons and on social costs.

Instead of considering marginal social changes in a policy, it is sometimes informative to know the gross gain accruing to the economy from the existence of a program at a level z compared to the alternative of shutting it down. This entails a comparison across programs and is what the parameter "treatment on the treated" is designed to estimate. The appropriate criterion for an all or nothing evaluation of a policy at level $Z = z$ is

$$\begin{aligned} A(z) = \{&\Pr(D = 1|Z = z)E(Y_1|D = 1, Z = z) \\ &+ \Pr(D = 0|Z = z)E(Y_0|D = 0, Z = z)\} \\ &- c(z) - E(Y_0|Z = 0)\end{aligned}$$

where $Z = 0$ corresponds to the case where there is no program and I assume $c(0) = 0$. If $A(z) > 0$, total output is increased by establishing the program at level z. Observe that when $Z = 0$, no one participates in the program, because it is not available. A crucial simplifying economic assumption makes it possible to use outcomes in the no treatment state for a policy with $Z = z$ to identify outcomes in the no program state ($Z = 0$). The assumption is the absence of general equilibrium effects.

Define $D(z)$ to be the conditional random variable D given $Z = z \neq 0$. In the special case where the outcome in the benchmark no program state 0 is the same whether or not the program exists, the absence of general equilibrium effects assumption entails that

(13a) $E(Y_0|D(z) = 0, Z = z) = E(Y_0|D(z) = 0, Z = 0)$

and

(13b) $E(Y_0|D(z) = 1, Z = z) = E(Y_0|D(z) = 1, Z = 0)$.

These expressions are defined, respectively, for persons who would not have participated ((13a) for $D(z) = 0$), or would have participated ((13b) for $D(z) = 1$), in the program if it existed at level $Z = z$. Equation (13a) states that the mean no treatment outcome for nonparticipants in the state when there is a program (at level $Z = z$) is the same as the mean no program outcome for the same group of persons who are nonparticipants in the program when it exists. Equation (13b) is a parallel assumption for participants in treatment when the program exists at level $Z = z$. It asserts that their mean no treatment outcome under the regime $Z = z$ would have been what their mean no program outcome would be under the no policy regime $Z = 0$ would be. $D(z)$ identifies participants or nonparticipants in the program $(Z = z)$ regime. The second conditioning argument on the right hand side of (13a) and (13b) allows for the possibility that the level of the effect depends on whether $Z = z$ or $Z = 0$, although this dependence is ruled out by virtue of assumptions (13a) and (13b).

Using the law of iterated expectations, and the assumption that participation $D(z)$ depends only on $Z = z$,

$$E(Y_0|Z = 0) = E(Y_0|D(z) = 1, Z = 0)\Pr(D(z) = 1) + E(Y_0|D(z) = 0, Z = 0)\Pr(D(z) = 0).$$

Using the assumed absence of general equilibrium effects as embodied in (13a) and (13b), we obtain

$$(14) \quad E(Y_0|Z = 0) = E(Y_0|D = 1, Z = z)\Pr(D = 1|Z = z) + E(Y_0|D = 0, Z = z)\Pr(D = 0|Z = z).$$

This decomposes $E(Y_0|Z = 0)$ into components attributable to persons who would have participated in the program if it were available and components attributable to those persons who would not have participated in the program if it were available at level $Z = z$.

Using (14) to substitute for $E(Y_0|Z = 0)$ in the definition of $A(z)$, we obtain

$$(15) \quad A(z) = \Pr(D = 1|Z = z)E(Y_1 - Y_0|D = 1, Z = z) - c(z).$$

If redistribution is costless, the output-maximizing solution also maximizes social welfare. For this important case, which is applicable to small-scale social programs with partial participation, the measure "treatment on the treated" is vindicated provided that additional data on costs are collected. If Z only affects participation but not outcomes, the traditional measure is fully vindicated because $E(y_1 - y_0|D = 1, Z = z) = E(y_1 - y_0|D = 1)$. For evaluating the effect of "fine-tuning" existing programs, measure $\Delta(z)$ is more appropriate. All conventional evaluation parameters abstract from cost effects of the policies being evaluated, however, and this is a serious limitation of them.

Note that in principle none of these policy parameters require micro data for their estimation. Knowing the impact of the program being evaluated on GNP would suffice. Only the well known limitations of using aggregate time series data to evaluate the impact of a small scale program prevents one from using them to identify the impact of the program, and thus avoid the problems of selection bias in microdata. (See Heckman and Smith 1997, for further discussion.)

The effect of randomly assigning persons in the population at large to the program is rarely of interest, except when coverage is universal and in that case $E(\Delta|X) = E(\Delta|X, D = 1)$. For that reason we do not discuss this parameter further in this section.

VII. Extensions To The Case of Continuous Treatment[13]

The previous analysis applies to the case where "treatment" is a discrete variable. I now briefly consider the case where treatment is a continuous variable. For specificity, consider a prototypical model of the earnings-schooling relationship patterned along the lines of Mincer (1974). Let y be earnings and S be schooling, then

$$\ln y = \alpha_0 + \alpha_1 S + V$$

where $E(V) = 0$ and α_1 may vary among people. As in the preceding sections, I implicitly condition on a set of variables X. The cost of schooling consists of foregone earnings while in school plus direct costs $c(S, Z, \eta)$ where Z denotes observed determinants of c and η represents unobserved determinants of c. To represent the variability in α_1, write

$$\alpha_1 = \bar{\alpha}_1 + \epsilon, \quad E(\epsilon) = 0.$$

Assuming that persons face discount rate $r > 0$ (which may vary among persons), and that they seek to maximize the present value of their earnings over an infinite horizon, the optimal level of schooling is the solution to

$$\underset{S}{\text{Max}}\, e^{-rS}\left[\frac{e^{\alpha_0+\alpha_1 S+V}}{r}\right] - c(S, Z, \eta)$$

where the term in brackets is the present value of schooling S at the end of schooling, e^{-rS} discounts schooling costs back to the beginning of life and $c(S, Z, \eta)$ is the direct cost of schooling (tuition or books). The necessary condition for optimal years of schooling is

$$\frac{(\alpha_1 - r)}{r} e^{\alpha_0+(\alpha_1-r)S+V} = \frac{\partial c}{\partial S}(S; Z; \eta).^{14}$$

For a positive solution, it is required that $\alpha_1 - r > 0$ if costs of school increase with S.

The solution to this problem for the case where

$$c(S; Z; \eta) = e^{\beta S+\gamma Z+\eta}, \quad \beta > 0$$

is particularly straightforward and I use this specification to make some simple points. The solution is

$$(16) \quad S = \left(\frac{1}{\alpha_1 - r - \beta}\right)[\gamma Z + \eta - V + \varphi]$$

13. *LATE* is not defined for continuous treatments. Thus, I do not discuss it in this context.

14. Second order sufficient conditions are

$$\frac{(\alpha_1 - r)^2}{r} e^{\alpha_0+(\alpha_1-r)S+V} - \frac{\partial^2 c}{\partial S^2}(S; Z; \eta) < 0.$$

Direct costs have to be rising sufficiently fast to guarantee an interior solution to this problem.

where

$$\varphi = -\left(\alpha_0 + \ln\left(\frac{\alpha_1 - r}{\beta r}\right)\right).$$

In this analysis β may vary among persons. To satisfy the second order condition for an optimum the additional condition, $\alpha_1 - r - \beta < 0$ must be satisfied. Thus $0 < \alpha_1 - r < \beta$ is a restriction on α_1, r and β required to make the model economically meaningful.

In this model, the "treatment" is S and for $S = s^*$, the treated are implicitly defined by (16). Thus persons who are treated at level $S = s^*$ are those persons with characteristics Z, η, V, α_1, r, and β such that (16) is satisfied with $S = s^*$. The mean effect of treatment on the treated for persons at schooling level s^* is

$$(17)\quad E\left(\frac{\partial \ln y}{\partial S} \,\middle|\, S = s^*\right)$$

$$= E\left(\alpha_1 | s^* = \left(\frac{1}{\alpha_1 - r - \beta}\right)(\gamma Z + \eta - V + \varphi), 0 < \alpha_1 - r < \beta\right).$$

If agents know only $\bar{\alpha}_1$ and not α_1 when they make their schooling decisions, and if they know V and η, then

$$E\left(\frac{\partial \ln y}{\partial S} \,\middle|\, S = s^*\right) = \bar{\alpha}_1,$$

which is the same parameter if α_1 is a constant for everyone. In this case, the second order condition is $0 < \bar{\alpha}_1 < \beta + r$. In estimating $\bar{\alpha}_1$ in this case, least squares may still be inconsistent, because of the dependence between V and S. This can arise because of the dependence between V and the unobservables in (16) which include V. In this context, Z is a valid instrumental variable provided that $E(V|Z) = 0$. If α_1 is random and forecastable, at least in part, by Z, the instrument breaks down for estimating parameter (17) because

$$E(\alpha_1 | S^* = s, 0 < \alpha_1 - r < \beta, Z = z)$$

depends on z.

When α_1 is forecastable, the parameter corresponding to the discrete outcome parameter $E(\Delta|X)$ in the continuous case is $E(\alpha_1 | 0 < \alpha_1 - r < \beta)$, the effect of picking someone at random from the population, and giving them an exogenous dose of schooling. This parameter does not answer a well-posed economic question nor does it equal $\bar{\alpha}_1$ unless α_1 is a constant. The instrumental variables estimator is inconsistent for the parameter in the general case when α_1 is known at the time schooling decisions are made because $E(\epsilon | S = s; Z = z) \neq 0$, and so the instrumental variable condition $E(\epsilon S | Z = z) = 0$ is not satisfied. If α_1 is unknown and people act on $\bar{\alpha}_1$, then the parameter corresponding to $E(\Delta|X)$ is $E(\alpha_1) = \bar{\alpha}_1$. In this case, instrumental variables is a consistent estimator of $\bar{\alpha}_1$ if $E(V|Z) = 0$. Assumptions about the information available to the agent play a crucial role in defining the economically interpretable parameters and justifying the application of instrumental variables to estimate them.

VIII. Summary

Statistical assumptions made in evaluation research are based on strong behavioral assumptions even though they are often disguised. This paper exposits how the method of instrumental variables that is widely used to estimate the impact of treatment on the treated or the impact of treatment on randomly assigned persons is based on the assumption (a) that persons with a given set of observed characteristics X respond identically to treatment or the assumption (b) that if responses conditional on X are heterogeneous, persons do not make their decisions to participate in the program based on unobserved (by the analyst) components of program gains. This latter assumption implies a strong form of ignorance or irrationality about unobserved components of gain on the part of the people being studied. It also implies that persons do not have private information that is useful in forecasting the gains that they use in making their decisions but that is not available to the analyst. If these implications are incorrect, the method of instrumental variables is inconsistent for estimating the effect of treatment on the treated. We have also considered the analysis of the *LATE* parameter of Imbens and Angrist. It is instrument dependent and is defined on a hypothetical, unobservable population. For policy interventions that only induce some persons to switch participation status and have no effects on nonswitchers and have no direct social costs, however, a version of *LATE* produces an economically interpretable parameter—the effect of a marginal policy change on per capita income.

A parallel analysis is presented for a model of the effect of schooling on earnings in which "treatments" are continuous. The same distinctions arise in this model, except the effect of picking someone at random and giving them schooling depends on whether the marginal treatment effect is variable or not, and if it is variable, whether or not it is anticipated, at least in part, when decisions to participate in the program are made.

Many methods besides instrumental variable methods are available for answering the evaluation questions considered in this paper. These methods do not rely on the strong behavioral assumptions required for the correct application of instrumental variables in the case when response to treatment is heterogeneous. See Bjorklund and Moffitt (1987), Heckman and Robb (1985, 1986), Heckman et al. (1994, 1996), Heckman, Ichimura, and Todd (1997a, b), and Heckman and Smith (1996) for discussion of these estimators and their properties, and applications to the evaluation of training programs.

References

Angrist, Joshua. 1990. "Lifetime Earnings and The Vietnam Era Draft Lottery: Evidence From Social Security Administration Records." *American Economic Review* 80(3): 313–35.

Ashenfelter, Orley, 1978. "Estimating The Effect of Training Programs on Earnings." *Review of Economics and Statistics* 60(1):47–57.

Ashenfelter, Orley, and David Card. 1985. "Using The Longitudinal Structure of Earnings To Estimate The Effect of Training Programs." *Review of Economics and Statistics* 67(3):648–60.

Bjorklund, Anders, and Robert Moffitt. 1987. "Estimation of Wage Gains and Welfare Gains in Self-Selection Models." *Review of Economics and Statistics* 69(1): 42–49.

Heckman, James. 1978. "Dummy Endogenous Variables in A Simultaneous Equations System." *Econometrica* 46(4):931–60.

———. 1990. "Varieties of Selection Bias." *American Economic Review* 80(2): 313–18.

———. 1996. "Randomization As An Instrumental Variable." *Review of Economics and Statistics* 77(2):336–41.

Heckman, James, and Bo Honoré. 1990. "Empirical Content of The Roy Model." *Econometrica* 58(5):1121–49.

Heckman, James, Hidehiko Ichimura, Jeffrey Smith, and Petra Todd. 1994. "Characterizing Selection Bias Using Experimental Data." *Econometrica*. Under revision.

———. 1996. "Sources of Selection Bias in Evaluating Programs: An Interpretation of Conventional Measures and Evidence on The Effectiveness of Matching As A Program Evaluation Method." *Proceedings of The National Academy of Sciences* 93(23): 13416–20.

Heckman, James, Hidehiko Ichimura, and Petra Todd. 1997a. "Matching As An Econometric Evaluation Estimator: Theory and Methods." *Review of Economic Studies*. Forthcoming.

———. 1997b. "Matching As An Econometric Estimator: Evidence from Evaluating a Job Training Program." *Review of Economic Studies*. Forthcoming.

Heckman, James, and Richard Robb. 1985. "Alternative Methods For Evaluating The Impact of Interventions." In *Longitudinal Analysis of Labor Market Data*, ed. James Heckman and Burton Singer, 156–245. New York: Wiley.

———. 1986. "Alternative Methods For Solving The Problem of Selection Bias in Evaluating The Impact of Treatments on Outcomes." In *Drawing Inferences From Self-Selected Samples*, ed. H. Wainer, 63–107. Berlin: Springer Verlag.

Heckman, James, and Guilherme Sedlacek. 1985. "Heterogeneity, Aggregation and Market Wage Functions: An Empirical Model of Self-Selection in the Labor Market." *Journal of Political Economy* 93(6):1077–25.

Heckman, James, and Jeffrey Smith. 1995. "Assessing the Case For Social Experiments." *Journal of Economic Perspectives* 9(2):85–100.

———. 1993. "Assessing The Case For Randomized Evaluation of Social Programs." In *Measuring Labour Market Measures*, ed. Karsten Jensen and Per Kongshoi Madsen, 35–95. Copenhagen, Denmark: Ministry of Labour.

———. 1996. "Experimental and Nonexperimental Evaluation." In *International Handbook of Labor Market Policy and Evaluation*, ed. G. Schmidt, J. O. Reilly and K. Schömann, 37–88. Cheltenham, U.K.: Elgar Publishers.

———. 1997. "Evaluating The Welfare State." In *The Ragnar Frisch Centenary*, ed. Steinar Strom. Econometric Society Monographs. Cambridge: Cambridge University Press.

Heckman, James, Jeffrey Smith, and Nancy Clements. 1997. "Making The Most Out of Program Evaluations and Social Experiments." *Review of Economic Studies*. Forthcoming.

Imbens, Guido, and Joshua Angrist. 1994. "Identification and Estimation of Local Average Treatment Effects." *Econometrica* 62(4):467–76.

LaLonde, Robert. 1986. "Evaluating The Econometric Evaluation of Training Programs with Experimental Data." *American Economic Review* 76(4):604–20.

Mincer, Jacob. 1974. *Schooling, Experience and Earnings*. Columbia University Press, New York.

Quandt, Richard. 1972. "Methods For Estimating Switching Regressions." *Journal of The American Statistical Association* 67(338):306–10.

———. 1988. *The Econometrics of Disequilibrium*. Oxford: Blackwell.

Robinson, Chris. 1989. "The Joint Determination of Union Status and Union Wage Effects: Some Tests of Alternative Models." *Journal of Political Economy* 97(3): 639–67.

Royden, H. L. 1968. *Real Analysis*, 2nd Edition. MacMillan: New York.

LINKED CITATIONS

You have printed the following article:

Instrumental Variables: A Study of Implicit Behavioral Assumptions Used in Making Program Evaluations
James Heckman
The Journal of Human Resources, Vol. 32, No. 3. (Summer, 1997), pp. 441-462.
Stable URL:
http://links.jstor.org/sici?sici=0022-166X%28199722%2932%3A3%3C441%3AIVASOI%3E2.0.CO%3B2-P

[Footnotes]

[1] **Randomization as an Instrumental Variable**
James J. Heckman
The Review of Economics and Statistics, Vol. 78, No. 2. (May, 1996), pp. 336-341.
Stable URL:
http://links.jstor.org/sici?sici=0034-6535%28199605%2978%3A2%3C336%3ARAAIV%3E2.0.CO%3B2-V

[11] **Heterogeneity, Aggregation, and Market Wage Functions: An Empirical Model of Self-Selection in the Labor Market**
James J. Heckman; Guilherme Sedlacek
The Journal of Political Economy, Vol. 93, No. 6. (Dec., 1985), pp. 1077-1125.
Stable URL:
http://links.jstor.org/sici?sici=0022-3808%28198512%2993%3A6%3C1077%3AHAAMWF%3E2.0.CO%3B2-P

References

Lifetime Earnings and the Vietnam Era Draft Lottery: Evidence from Social Security Administrative Records
Joshua D. Angrist
The American Economic Review, Vol. 80, No. 3. (Jun., 1990), pp. 313-336.
Stable URL:
http://links.jstor.org/sici?sici=0002-8282%28199006%2980%3A3%3C313%3ALEATVE%3E2.0.CO%3B2-H

NOTE: *The reference numbering from the original has been maintained in this citation list.*

LINKED CITATIONS

Estimating the Effect of Training Programs on Earnings
Orley Ashenfelter
The Review of Economics and Statistics, Vol. 60, No. 1. (Feb., 1978), pp. 47-57.
Stable URL:
http://links.jstor.org/sici?sici=0034-6535%28197802%2960%3A1%3C47%3AETEOTP%3E2.0.CO%3B2-K

Using the Longitudinal Structure of Earnings to Estimate the Effect of Training Programs
Orley Ashenfelter; David Card
The Review of Economics and Statistics, Vol. 67, No. 4. (Nov., 1985), pp. 648-660.
Stable URL:
http://links.jstor.org/sici?sici=0034-6535%28198511%2967%3A4%3C648%3AUTLSOE%3E2.0.CO%3B2-N

The Estimation of Wage Gains and Welfare Gains in Self-Selection Models
Anders Björklund; Robert Moffitt
The Review of Economics and Statistics, Vol. 69, No. 1. (Feb., 1987), pp. 42-49.
Stable URL:
http://links.jstor.org/sici?sici=0034-6535%28198702%2969%3A1%3C42%3ATEOWGA%3E2.0.CO%3B2-D

Dummy Endogenous Variables in a Simultaneous Equation System
James J. Heckman
Econometrica, Vol. 46, No. 4. (Jul., 1978), pp. 931-959.
Stable URL:
http://links.jstor.org/sici?sici=0012-9682%28197807%2946%3A4%3C931%3ADEVIAS%3E2.0.CO%3B2-F

Varieties of Selection Bias
James Heckman
The American Economic Review, Vol. 80, No. 2, Papers and Proceedings of the Hundred and Second Annual Meeting of the American Economic Association. (May, 1990), pp. 313-318.
Stable URL:
http://links.jstor.org/sici?sici=0002-8282%28199005%2980%3A2%3C313%3AVOSB%3E2.0.CO%3B2-G

Randomization as an Instrumental Variable
James J. Heckman
The Review of Economics and Statistics, Vol. 78, No. 2. (May, 1996), pp. 336-341.
Stable URL:
http://links.jstor.org/sici?sici=0034-6535%28199605%2978%3A2%3C336%3ARAAIV%3E2.0.CO%3B2-V

NOTE: *The reference numbering from the original has been maintained in this citation list.*

LINKED CITATIONS

The Empirical Content of the Roy Model
James J. Heckman; Bo E. Honoré
Econometrica, Vol. 58, No. 5. (Sep., 1990), pp. 1121-1149.
Stable URL:
http://links.jstor.org/sici?sici=0012-9682%28199009%2958%3A5%3C1121%3ATECOTR%3E2.0.CO%3B2-K

Heterogeneity, Aggregation, and Market Wage Functions: An Empirical Model of Self-Selection in the Labor Market
James J. Heckman; Guilherme Sedlacek
The Journal of Political Economy, Vol. 93, No. 6. (Dec., 1985), pp. 1077-1125.
Stable URL:
http://links.jstor.org/sici?sici=0022-3808%28198512%2993%3A6%3C1077%3AHAAMWF%3E2.0.CO%3B2-P

Assessing the Case for Social Experiments
James J. Heckman; Jeffrey A. Smith
The Journal of Economic Perspectives, Vol. 9, No. 2. (Spring, 1995), pp. 85-110.
Stable URL:
http://links.jstor.org/sici?sici=0895-3309%28199521%299%3A2%3C85%3AATCFSE%3E2.0.CO%3B2-K

Identification and Estimation of Local Average Treatment Effects
Guido W. Imbens; Joshua D. Angrist
Econometrica, Vol. 62, No. 2. (Mar., 1994), pp. 467-475.
Stable URL:
http://links.jstor.org/sici?sici=0012-9682%28199403%2962%3A2%3C467%3AIAEOLA%3E2.0.CO%3B2-2

Evaluating the Econometric Evaluations of Training Programs with Experimental Data
Robert J. LaLonde
The American Economic Review, Vol. 76, No. 4. (Sep., 1986), pp. 604-620.
Stable URL:
http://links.jstor.org/sici?sici=0002-8282%28198609%2976%3A4%3C604%3AETEEOT%3E2.0.CO%3B2-P

A New Approach to Estimating Switching Regressions
Richard E. Quandt
Journal of the American Statistical Association, Vol. 67, No. 338. (Jun., 1972), pp. 306-310.
Stable URL:
http://links.jstor.org/sici?sici=0162-1459%28197206%2967%3A338%3C306%3AANATES%3E2.0.CO%3B2-3

NOTE: *The reference numbering from the original has been maintained in this citation list.*

LINKED CITATIONS

The Joint Determination of Union Status and Union Wage Effects: Some Tests of Alternative Models
Chris Robinson
The Journal of Political Economy, Vol. 97, No. 3. (Jun., 1989), pp. 639-667.
Stable URL:
http://links.jstor.org/sici?sici=0022-3808%28198906%2997%3A3%3C639%3ATJDOUS%3E2.0.CO%3B2-D

NOTE: *The reference numbering from the original has been maintained in this citation list.*