TECHNICAL WORKING PAPER SERIES

INSTRUMENTAL VARIABLES:
A CAUTIONARY TALE

James J. Heckman

Technical Working Paper No. 185

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
September 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 his comments. This paper is part of NBER's research program in Labor Studies. Any opinions expressed are those of the author and not those of the National Bureau of Economic Research.

NBER Technical Working Paper #185
September 1995

INSTRUMENTAL VARIABLES: A CAUTIONARY TALE

ABSTRACT

This paper considers the use of instrumental variables to estimate the mean effect of treatment on the treated. It reviews previous work on this topic by Heckman and Robb (1985, 1986) and demonstrates that (a) unless the effect of treatment is the same for everyone (conditional on observables), or (b) treatment effects are variable across persons but the person-specific component of the variability not forecastable by observables does not determine participation in the program, widely-used instrumental variable methods produce inconsistent estimators of the parameter of interest. Neither assumption is very palatable. The first assumes a homogeneity that is implausible. The second assumes either very rich data available to the econometrician or that the persons being studied either do not have better information than the econometrician or that they do not use it. Instrumental variable methods do not provide a general solution to the evaluation problem.

James J. Heckman
Department of Economics
University of Chicago
1126 E. 59th Street
Chicago, IL 60637
and NBER

In the last years of his life, Sar Levitan and I had a series of conversations about alternative approaches for evaluating social programs. They began with a chance encounter in Washington, D.C. in 1991 and continued until the end of his life.

Our conversations focused on questions central to the field of evaluation research. We both felt that widely-used statistical models made strong implicit assumptions about behavior which, when exposited, would make them less attractive. This paper summarizes our discussion of the widely-used method of instrumental variables.

The method of instrumental variables is now widely used in evaluation research. It has recently been promoted by Angrist, Imbens and Rubin (1996). To understand the method, consider the following commonly-used model for evaluating programs. Let Y be an outcome such as earnings. Let $D = 1$ if a person participates in a training program, $D = 0$ otherwise. Using a standard regression model, write

$$Y = \alpha + \beta D + U$$

where U is an error term with mean zero ($E(U) = 0$). This term arises from omitted variables that may determine the outcome.

Much concern has been expressed about the possibility that D is correlated with U. Persons who would have high values of U(or Y) in the absence of the program may be the ones who go into it. This could happen if there is "cream-skimming" in which program officials select persons with high U. Formally, the fear translates into the concern that

$$E(U \mid D = 1) \neq 0.$$

In our example, this term would be positive. Then simple regression estimates of β are upward biased.

An instrumental variable Z has two properties. It is uncorrelated with U:

(a) $E(U \mid Z) = 0.$

It is correlated with D:

(b) $E(D \mid Z) = \Pr(D = 1 \mid Z)$ is a non-trivial function of Z taking at least two distinct values for two distinct values of Z.

Applying these properties to the outcome equation, take expectations conditional on Z:

$$E(Y \mid Z) = \alpha + \beta \Pr(D = 1 \mid Z).$$

Using the assumption that there are two distinct values of Z, say Z_1 and Z_2, with distinct probability values (i.e. distinct values of Pr(D=1| X)) we can solve for β.

$$E(Y \mid Z_1) = \alpha + \beta \Pr(D = 1 \mid Z_1)$$
$$E(Y \mid Z_2) = \alpha + \beta \Pr(D = 1 \mid Z_2)$$

$$\beta = \frac{E(Y \mid Z_1) - E(Y \mid Z_2)}{\Pr(D = 1 \mid Z_1) - \Pr(D = 1 \mid Z_2)}.$$

Using sample moments to replace population moments, we obtain the instrumental variables estimator.

The method requires an instrument Z. How credible are the instruments? Under what conditions does the method work? Can the method be generalized to cover the plausible case where β differs among people *i.e.*, there is variable response to treatment instead of a common response for everyone?

This paper addresses these questions. I establish that in cases in which the responses to the treatment vary, the instrumental variable argument fails unless person-specific responses to treatment do not influence the decision to participate in the program. This requires that individual gains from the program that cannot be predicted from variables available to observing social scientists do not influence the program participation decision. In the likely case in which individuals possess private information about gains from the program that cannot be proxied by the available data, instrumental variables methods do not estimate economically interesting evaluation parameters. Instrumental variable models are extremely sensitive to assumptions about how people process information.

In order to establish these points it is necessary to step back from the standard regression model and consider a more general model that captures the notion of heterogenous responses.

1. The Evaluation Problem

A person has two potential states, only one of which is realized. 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 is to replace the missing data on persons using group means or some other group statistics.

Using group statistics, what are the parameters of interest for evaluating social programs? Many have been proposed. It is of interest to explore the impacts of programs on distributions of outcomes. This is done in other papers. (See Clements, Heckman and Smith, 1993, and Heckman and Smith, 1995). Most 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 is

(1) $$\mathrm{E}(\Delta \mid D=1) = E(Y^1 - Y^0 \mid D=1).$$

(These expectations can be defined conditionally on explanatory variables X so different groups have different outcome measures.) We know or can reliably estimate $\mathrm{E}(Y^1 \mid D = 1)$. This is what participants experience. We don't know $\mathrm{E}(Y^0 \mid D = 1)$ — what participants would have experienced had they not participated. This is the missing counterfactual.

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) $$\mathrm{E}(\Delta) = E(Y^1 - Y^0).$$

This intuitively appealing counterfactual is very difficult to estimate. Picking a millionaire at random to participate in a training program for low skilled workers may be an intriguing thought experiment but is neither policy relevant nor feasible. It is 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.

Understanding the difference between these two counterfactuals is of central importance in understanding competing approaches followed in the evaluation literature. Conventional practice in the econometric evaluation community (*e.g.*

Ashenfelter (1978), LaLonde (1986), Ashenfelter and Card (1985)) assumes a special model in which these two counterfactuals are the same.

2. Constructing Counterfactuals

How does one go about constructing counterfactuals? That is the topic of a vast literature. Much of the literature on constructing counterfactuals draws on econometric models.

Consider a model in which the outcomes depend on explanatory variables X. In the traditional regression setting, $Y^0 = X\beta^0 + U^0$ and $Y^1 = X\beta^1 + U^1$ where $E(U^0 \mid X) = 0$ and $E(U^1 \mid X) = 0$. Nothing said here relies on linear regression models. We can just as easily use more general models. These more general models are really just models for conditional means. From now on, I will work with means for a particular group with characteristics X. $\mu^1(X)$ is $E(Y^1 \mid X)$. $\mu^0(X)$ is $E(Y^0 \mid X)$.

(3a) $$Y^0 = \mu^0(X) + U^0$$

(3b) $$Y^1 = \mu^1(X) + U^1$$

where $E(U^0 \mid X) = 0$ and $E(U^1 \mid X) = 0$. In the regression setting, $\mu^0(X) = X\beta^0$ and $\mu^1(X) = X\beta^1$. Observed outcome Y can be written as

$$\mathrm{Y} = \mathrm{DY}^1 + (1 - \mathrm{D})\mathrm{Y}^0$$

i.e. it is either Y^1 or Y^0. If we insert (3a) and (3b) into this expression,

(4) $$\mathrm{Y} = \mu^0(X) + D(\mu^1(X) - \mu^0(X) + U^1 - U^0) + U^0.$$

This is a "two regime" or "switching regression" model. (See Quandt, 1972). Labor economists call it a Roy model (see, *e.g.* Willis and Rosen, 1979, and Heckman and Honoré, 1990).

The term multiplying D is the gain. The gain has two components: $\mu^1(X) - \mu^0(X)$ — the gain for the average person — and $U^1 - U^0$ — the idiosyncratic gain for a person who differs from the average. In this notation, then, the average gain is

$$E(\Delta \mid X) = \mu^1(X) - \mu^0(X).$$

The effect of treatment on the treated is

$$E(\Delta \mid X, D = 1) = \mu^1(X) - \mu^0(X) + E(U^1 - U^0 \mid X, D = 1).$$

The latter expression differs from the former by the additional term $E(U^1 - U^0 \mid X, D = 1)$. This tells you how much the *gain of participants* differs from the average gain that would be experienced by the entire population. This is the gain to the movers from going "0" to "1." Even though on average, persons gain $\mu^1(X) - \mu^0(X)$, for *participants* the gain will be different. We may rewrite equation (4) in terms of these parameters

(5) $\quad Y = \mu^0(X) + D[E(\Delta \mid X)] + \{U^0 + D(U^1 - U^0)\}$

and

(6) $Y = \mu^0(X) + D[E(\Delta \mid X, D = 1)] + \{U^0 + D[U^1 - U^0 - E(U^1 - U^0 | X, D = 1)\}.$

This notation is dense. It can be simplified back to the notation used in the introduction. Let $\mu^0(X) = \alpha(X)$, $\beta(X) = \mu^1(X) - \mu^0(X) + U^1 - U^0$ and let the mean of β conditional on X be $E(\beta(X) \mid X) = \overline{\beta}(X)$. Let $\varepsilon = U^1 - U^0$. Let $U^0 = U$. For notational simplicity, the dependence of α and β on X is left implicit: $\alpha(X) = \alpha, \beta(X) = \beta$. Then (4) can be written as a dummy-variable regression, like we encountered before:

(4′) $\quad \text{Y} = \alpha + D\beta + U.$

Now, however, even after conditioning on X, β varies in the population reflecting individual heterogeneity in response to treatment.

Equation (5) can be written as

(5′) $\quad Y = \alpha + D\overline{\beta} + \{U + D\varepsilon\}.$

$\overline{\beta}$ is the effect of placing an average person in the population at large into the program. (In general, $\bar{\beta}$ is a function of X, $\bar{\beta}(X)$. Again the conditioning is left implicit). The effect of treatment on the treated is

$$E(\Delta \mid X, D=1) = E(\beta \mid D=1, X) = \overline{\beta} + E(\varepsilon \mid X, D=1) = \beta^{*}.$$

Again, I leave implicit the dependence of β^{*} on X. X plays no important role here because it is assumed to be mean independent of $U, E(U \mid X) = 0$. The analogue to equation (6) is

(6′) $$Y = \alpha + D\beta^{*} + \{U + D(\varepsilon - E(\varepsilon \mid D=1))\}.$$

Can we run a regression of Y on D with an intercept to estimate $\overline{\beta}$ and β^{*}? The regression coefficient of D can always be written as the difference between two means. Let $\widetilde{\beta}$ be the probability limit — the value of the regression coefficient in large samples under conventional assumptions. There are three alternative representations of this limit.

(7a) $$\widetilde{\beta} = E(Y \mid X, D=1) - E(Y \mid X, D=0)$$

(7b) $$\widetilde{\beta} = \overline{\beta} + E(\varepsilon \mid X, D=1) + [E(U \mid X, D=1) - E(U \mid X, D=0)]$$

(7c) $$\widetilde{\beta} = \beta^{*} + E[U \mid X, D=1] - E[U \mid X, D=0].$$

From (7b), we see that $\widetilde{\beta}$ is biased for $\overline{\beta}$ by an amount

$$E(\varepsilon \mid X, D=1) + E(U \mid X, D=1) - E(U \mid X, D=0).$$

From (7c), we see that $\widetilde{\beta}$ is biased for β^{*} by an amount

(8) $$E(U \mid X, D=1) - E(U \mid X, D=0).$$

This term (8) is sometimes called the selection bias term. It tells us how the outcome in the base state differs between program participants and nonparticipants. Such differences cannot be attributed to the program. In terms of the previous notation, recalling that $U^0 = U$,

$$E(U \mid X, D=1) - E(U \mid X, D=0) = E(U^0 \mid X, D=1) - E(U^0 \mid X, D=0)$$

$$= E(Y^0 \mid X, D = 1) - E(Y^0 \mid X, D = 0).$$

The difference between $\overline{\beta}$ and β^* is the difference between the unobservable gain for an average person in the population (defined to be zero) and the unobservable gain for the average participant:

$$E(\varepsilon \mid X, D = 1) = E(U^1 - U^0 \mid X, D = 1).$$

β^* differs from $\overline{\beta}$ by the amount of the gain in unobservables between the two groups $(D = 1, D = 0)$. These unobservables may be observed by the person or persons deciding to go into the program. They are unobserved only from the point of view of the social scientist trying to estimate the impact of the program. They coincide when the mean gain in the unobservable conditional on D is zero i.e.

$$E(\varepsilon \mid X, D = 1) = 0 = E(U^1 - U^0 \mid X, D = 1).$$

We now present two special cases when $\overline{\beta} = \beta^*$.

3. When Does $\overline{\beta} = \beta^*$?

There are two important special cases when $\overline{\beta} = \beta^*$. The first is when $U^1 = U^0$. In this case, there are no unobservable components of the gain. This model — called the "dummy endogenous variable model" (see, *e.g.* Heckman, 1978) — is widely used in applied work (see Ashenfelter, 1978, 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 is more subtle. $U^1 \neq U^0$. But $U^1 - U^0$ does not determine who goes into the program. Suppose, for example, that at the time people go into the program they do not know $\varepsilon = U^1 - U^0$. Their best forecast of ε is zero or some other constant. Then if their experience of ε is typical of that of the entire population, $E(\varepsilon \mid X, D=1)=0$, and $\overline{\beta}=\beta^*$.

This case shares many features in common with the "random coefficients" model of traditional econometrics. This comparison is pursued in the Appendix.

Observe that in either the case where $U^1 = U^0$ or the case where ε is not forecastable at the date enrollment decisions are made, the problem of estimating $\overline{\beta}$ or β^*using the difference in outcomes between participants and nonparticipants, comes down to the problem arising from D being correlated with, or stochastically dependent on, U.

Also note that for the estimation of (6′), from the definition of β^*, D is constructed to be uncorrelated with $D(\varepsilon - E(\varepsilon \mid X, D=1))$. Thus *irrespective of whether* $E(\varepsilon \mid X, D=1)=0$ or not

$$E(\varepsilon - E(\varepsilon \mid X, D=1) \mid X, D=1) = E(\varepsilon \mid X, D=1) - E(\varepsilon \mid X, D=1) = 0.$$

Thus, even in the case where participation in the program is made at least in part on unobservable components of gain, the only source of correlation between the "error term" and the treatment variable "D" arises from U, if the coefficient of interest is β^*, the effect of treatment on the treated.

4. The Method Of Instrumental Variables

A standard method for estimating parameters in econometrics is the method of instrumental variables discussed in the introduction for the conventional simultaneous equation model. Here we consider them in a more general context. Instrumental variables must satisfy two basic conditions and a third derived condition. They are mean-independent of the error terms of equations (5′) and (6′) *i.e.*

(A-1-a) $\quad E[U + D\,\varepsilon \mid X, Z] = 0$

for identifying $\overline{\beta}$ or

(A-1-b) $E[U + D(\varepsilon - E(\varepsilon \mid D = 1)) \mid X, Z] = 0$

for identifying β^*. A second condition is that D depends on Z in the following way:

(A-2) $E(D \mid X, Z) = \Pr(D = 1 \mid X, Z)$

is a function of Z. Z is not fully explained by X. Implicit is the idea that the probability is defined for two or more values of Z, so that the probability of participation is a nontrivial function of Z, and the values of the probability differ for different Zs at least for some X.

The important idea in this condition is that the probability of participation depends on Z as well as on X. A third condition is that the dependence of Y on Z operates only through D. This condition is really a consequence of the first two conditions, but its role is so central that I break it out as a separate condition. Reminding the reader that α, $\overline{\beta}$ and β^* may depend on X, for $\overline{\beta}$ the derived third assumption is

(A-3-a) $E(Y \mid X, Z) = \alpha + \overline{\beta} E(D \mid X, Z) = \alpha + \overline{\beta} \Pr(D = 1 \mid X, Z)$.

For β^* the assumption is

(A-3-b) $E(Y \mid X, Z) = \alpha + \beta^* E(D \mid X, Z) = \alpha + \beta^* \Pr(D = 1 \mid X, Z)$.

For two distinct values of Z, say Z_1 and Z_2, such that $\Pr(D = 1 \mid X, Z_1) \neq \Pr(D = 1 \mid X, Z_2)$, we can identify $\overline{\beta}$ or β^* by forming a simple mean difference. Thus from (A-3-a), (A-2) and (A-1-a),

$$E(Y \mid X, Z_1) - E(Y \mid X, Z_2) = \overline{\beta}[\Pr(D = 1 \mid X, Z_1) - \Pr(D = 1 \mid X, Z_2)]$$

so

$$\overline{\beta} = \frac{E(Y \mid X, Z_1) - E(Y \mid X, Z_2)}{\Pr(D = 1 \mid X, Z_1) - \Pr(D = 1 \mid X, Z_2)}.$$

Replacing population means with sample means produces the instrumental variable estimator which, under standard conditions, converges to $\widetilde{\beta}(X)$. By similar reasoning, using (A-3-b), (A-2) and (A-1-b),

$$\beta^* = \frac{E(Y \mid X, Z_1) - E(Y \mid X, Z_2)}{\Pr(D = 1 \mid X, Z_1) - \Pr(D = 1 \mid X, Z_2)}.$$

Loosely speaking, instruments are variables that "don't belong in the population outcome equation" but which "belong" in equation predicting program participation.

In the common effect model where $\varepsilon = 0$ or $U^1 - U^0 = 0$, (Case 1) Heckman (1978) showed that conventional instrumental variables methods would identify $\beta^* = \overline{\beta}$. One needs to find some variable or variables Z "uncorrelated with U" that affect "D" and are not already in the outcome equation (do not enter α and β or β^*respectively).

In the model where ε is not a determinant of D *i.e.* where

$$\Pr(D = 1 \mid X, Z, \varepsilon) = \Pr(D = 1 \mid X, Z),$$

or

$$\Pr(D = 1 \mid X, Z, Y^1 - Y^0) = \Pr(D = 1 \mid X, Z)$$

the same conditions have to be satisfied for Z. i.e. "uncorrelated with U" and "not in the outcome equation." We can ignore the component $D\varepsilon$ because

$$E(D\varepsilon \mid X, Z) = E(\varepsilon \mid X, Z, D = 1) \Pr(D = 1 \mid X, Z) = 0$$

since

$$E(\varepsilon \mid X, Z, D = 1) = 0.$$

Thus the two special cases where $\overline{\beta} = \beta^*$ are cases where we can use conventional text-book instrumental variable models. (Heckman and Robb (1985, 1986) develop both of these cases.)

What about more general cases? Consider equation (6)′ associated with β^*. Since the only source of dependence between the error term and D is through U and not $D(\varepsilon - E(\varepsilon \mid X, D = 1))$, the instrumental variable method looks promising. If assumptions (A-1-b), (A-2) and (A-3-b) are satisfied, the method can be used to identify β^*. What is required is a variable Z that affects participation but does not enter the parameters in the equation of interest. This requires that

$$\begin{aligned} E(\overline{\beta} + \varepsilon \mid X, Z, D &= 1) = \mu^1(x) - \mu^0(X) + E(\varepsilon \mid X, Z, D = 1) \\ &= \mu^1(X) - \mu^0(X) + E(U^1 - U^0 \mid X, Z, D = 1) \end{aligned}$$

not depend on Z.

This is mechanically satisfied in the first two cases where $U^1 - U^0 = 0$, (Case 1) or $E(\varepsilon \mid X, Z, D = 1) = 0$ (ε cannot be forecast by Z and X) (Case 2). Actually it is also satisfied in a third case where ε cannot be forecast by Z but can be forecast by X. But in general, this condition is very difficult to satisfy. If purposive persons select into the program on the basis of the gain in unobservables or on the basis of the variables that are (stochastically) dependent on the gain in unobservables, this condition will not be satisfied. (See Heckman and Robb, 1985, 1986.)

Any application of the method of instrumental variables for estimating the mean effect of "treatment on the treated" in the case where the response to "treatment" varies among persons requires that a behavioral assumption be made about how persons make their decisions about program participation. The issue cannot be settled by a statistical analysis.

Consider an example that is widely cited as a triumph of the method of instrumental variables. Draft lottery numbers are alleged to be ideal instrumental variables for identifying the effect of military service on earnings (Angrist, 1990). The 1969 lottery randomly assigned different priority numbers to persons of different birth dates. The higher the number, 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, $U^1 - U^0 = \varepsilon$, or base their decisions to go into the military on variables correlated with unobservable components ε, persons with high Z for whom $D = 1$ (they serve in the military) are likely to have high values of ε. This violates assumption (A-3-b), and makes the birth date number an invalid instrument. It is plausible that the persons who are deciding to go into the military have more information at their disposal than analysts using standard data sets. If this information is at all useful in predicting the gain from going into the military, the draft number is not a valid instrument.

The draft lottery number is a poor instrument for two other reasons. (1) Z is likely to be an X. Persons with high Z (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 (A-3-b) to be violated because Z is really an X. (2) Switching from a regime of a capricious draft to a lottery reduces uncertainty and is likely to change the investment behavior of persons of all levels of Z. In this instance, the switch from a draft to a lottery affects both β^* and $\overline{\beta}$ since it fundamentally alters schooling and job training investment decisions. Thus, knowing how military service affects earnings during the period of a lottery would not be informative on how military service affected earnings during the period of an ordinary draft.

As a second example, it is sometimes suggested that cross-state variation in welfare benefit schedules can be used as instrumental variables for estimating the effect of "treatment on the treated" for participants in training programs (see Moffitt, 1996). Suppose that Y^0 refers to the earnings of untrained lowskill persons. Y^1 is earnings in the trained state. Parameters of welfare benefit determinants, Z, do not plausibly enter $\mu^0(X)$ or $\mu^1(X)$. But they could enter

$$E(\varepsilon \mid X, Z, D = 1) = E(U^1 - U^0 \mid X, Z, D = 1)$$

because if the distribution of (U^0, U^1) is the same across states, and 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 participants enter the program with at least partial knowledge of $U^1 - U^0$. Then assumption (A-3-b) is violated, and cross-state variation in benefits do not identify β^* through the method of instrumental variables.

It might be thought that since

$$E[U^1 - U^0 - E(U^1 - U^0 \mid X, D = 1) \mid X, D = 1] = 0,$$

or

$$E(\varepsilon - E(\varepsilon \mid X, D = 1) \mid X, D = 1) = 0$$

it follows that $E[U^1 - U^0 - E(U^1 - U^0 \mid X, D = 1) \mid X, Z, D = 1] = 0$. This is not true. In general

$$E(U^1 - U^0 \mid X, Z, D = 1) \neq E(U^1 - U^0 \mid X, D = 1).$$

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

$$E(U \mid X, Z) = 0,$$

it does not follow that (A-1-b) is satisfied. Thus Z is not a good instrument for identifying β^*. For similar reasons, (A-1-a) is unlikely to be satisfied even if

$$E[U \mid X, Z] = 0$$

because

$$E[D\varepsilon \mid X, Z] = E[U^1 - U^0 \mid X, Z, D = 1] \Pr(D = 1 \mid X, Z)$$

does not equal zero. Thus Z is not a good instrument for identifying $\overline{\beta}$ either.

5. Summary

Sar Levitan and I both agreed that statistical assumptions often made in evaluation research are based on strong behavioral assumptions. This paper exposits how the widely used method of instrumental variables is based on the assumption (a) that persons respond identically to treatment or the assumption (b) that if responses are heterogeneous, persons do not make their decisions to participate in the program based on unobserved components of program gains. This assumption requires a strong form of ignorance about unobserved components of gain on the part of the people we study. It also implies that persons do not have private information useful in forecasting the gains that are not available to the analyst. If these assumptions are incorrect, the method of instrumental variables fails for estimating the effect of treatment on the treated.

Appendix: Comparison With Conventional Random Coefficient Models

The random coefficients model captures the idea that the response of a person to a variable may differ across persons. (See Judge et.al. (1980) for a reference to such models. Thus, in a traditional linear regression model

$$Y_i = X_i \gamma_i + U_i,$$

we explicitly allow the coefficient to bear a subscript i to identify variables that differ among persons. This is an intuitively attractive model because it allows for person-specific responses to changes in X.

The mean value of γ_i is $\bar{\gamma} = E(\gamma_i)$. We assume that this mean is finite. Assuming that X_i, γ_i and U_i are statistically independent of each other, and $\mathrm{E}(\mathrm{U}_i) = 0$, we may write $Y_i = X_i\bar{\gamma} + \{U_i + X_i\varepsilon_i\}$ where $\varepsilon_i = \gamma_i - \bar{\gamma}$, which is in the form of a common-coefficient model.

The composite error term has mean zero, and conditional on X_i, it has a variance component added to the usual error term. Thus

$$\text{Variance } (U_i + X_i\varepsilon_i \mid X_i) = Var(U_i) + Var(\varepsilon_i)X_i^2.$$

In this conventional model, X_i could be a dummy variable for "treatment" in which case we have a "components of variance" model. Persons receiving treatment have additional variability in outcomes due to their variability in the response to treatment.

The model in the text for the case of $\varepsilon = U^1 - U^0 \neq 0$ has D mean independent of $\varepsilon(E(\varepsilon \mid X, D = 1) = 0))$. But unlike the traditional random coefficient model, there is no presumption the error component that εD is (mean) independent of

U, or that D is (mean) independent of U. Thus there may be a selection problem ($E(U \mid X, D = 1) \neq 0$). However, if $E(\varepsilon \mid X, D = 1) = 0$, the only econometric problem in estimating the parameters of (5′) or (6′) arises from mean dependence between U and D, the first component of the error term in (5′) or (6′).

References

[1] Angrist, Joshua, "Lifetime Earnings and The Vietnam Era Draft Lottery: Evidence From Social Security Administration Records," *American Economic Review*, Vol. 80, 1990, 313-335.

[2] ________, Guido Imbens and Donald Rubin, "Identification of Causal Effects Using Instrumental Variables," forthcoming *Journal of the American Statistical Association*, 1996.

[3] Ashenfelter, Orley, "Estimating The Effect of Training Programs on Earnings," *Review of Economics and Statistics*, February, 1978, 47-57.

[4] Clements, Nancy, James Heckman and Jeffrey Smith, "Making The Most Out of Social Experiments: The Intrinsic Uncertainty in Evidence From Randomized Trials With An Application To The National JTPA Experiment", forthcoming, *Review of Economic Studies*, 1996.

[5] Heckman, James, "Dummy Endogenous Variables in A Simultaneous Equations System," *Econometrica*, July, 1978, 679-694.

[6] Heckman, James and Bo Honoré, "Empirical Content of The Roy Model," *Econometrica*, Vol. 58, September, 1990, 1121-1149.

[7] ________, and Richard Robb, "Alternative Methods For Evaluating The Impact of Interventions," in **Longitudinal Analysis of Labor Market Data**, New York, NY: Wiley, 1985.

[8] ________,and Richard Robb, "Alternative Methods For Solving The Problem of Selection Bias in Evaluating The Impact of Treatments on Outcomes,"

in H. Wainer, ed., **Drawing Inferences From Self-Selected Samples**, Berlin: Springer Verlag, 1986.

[9] Heckman, James and Jeffrey Smith, "Evaluating The Welfare State", Frisch Symposium, Oslo, Norway, March, 1995.

[10] Judge, George, William Griffiths, R. Carter Hill, Tsoung-Chao Lee, **The Theory and Practice of Econometrics**, New York, NY: John Wiley, 1980.

[11] LaLonde, Robert, "Evaluating The Econometric Evaluation of Training Programs," *American Economic Review*, Vol. 76, June, 1986, 604-620.

[12] Moffitt, Robert, "Comment on Angrist, et.al," forthcoming, *Journal of The American Statistical Association*, 1996.

[13] Quandt, Richard, "Methods For Estimating Switching Regressions," *Journal of The American Statistical Association*, 1972.

[14] Willis, Robert and Sherwin Rosen, "Education and Self-Selection," *Journal of Political Economy* (Supplement), Vol. 87, October, 1979, S7-S36.