NBER TECHNICAL WORKING PAPERS SERIES

IDENTIFICATION AND ESTIMATION OF LOCAL AVERAGE TREATMENT EFFECTS

Joshua D. Angrist

Guido W. Imbens

Technical Working Paper No. 118

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
December 1991

We are grateful for comments and suggestions by Gary Chamberlain, Larry Katz and Don Rubin. This paper is part of NBER's research program in Labor Studies. Any opinions expressed are those of the authors and not those of the National Bureau of Economic Research.

NBER Technical Working Paper #118
December 1991

IDENTIFICATION AND ESTIMATION OF LOCAL AVERAGE TREATMENT EFFECTS

ABSTRACT

We investigate conditions sufficient for identification of average treatment effects using instrumental variables. First we show that the existence of valid instruments is not sufficient to identify any meaningful average treatment effect. We then establish that the combination of an instrument and a condition on the relation between the instrument and the participation status is sufficient for identification of a local average treatment effect for those who can be induced to change their participation status by changing the value of the instrument. Finally we derive the probability limit of the standard IV estimator under these conditions. It is seen to be a weighted average of local average treatment effects.

Joshua D. Angrist
Department of Economics
Hebrew University
Mt. Scopus
Jerusalem, 91905
ISRAEL

NBER and Harvard University

Guido W. Imbens
Department of Economics
Harvard University
Cambridge, MA 02138

1. INTRODUCTION.

Most applied science is concerned with uncovering and verifying causal relationships. Therefore, many disciplines have spawned a literature concerned with estimating the effect of treatments, interventions, and programs. In the evaluation of new medical treatments and drugs, standard have emerged that researchers must usually follow for their work to be considered credible. Although the need for randomization is sometimes a subject of dispute (Royall, [1991]), random assigment of treatment and concurrent data collection on treatment and control groups is the norm in medical research . In contrast, the use of random assigment to evaluate social programs remains controversial. While some researchers have argued that evaluations based on observational studies are not credible (Lalonde [1986]), others argue that experiments can never be a complete substitute for analyzing observational data (Heckman and Hotz [1990]).

Disputes over the possibility and even desirability of randomization in social research have led researchers to search for reliable methods of estimating treatment effects from observational data. Recently, much research has been geared towards establishing conditions that guarantee identification of treatment effects without relying on functional form restrictions or distributional assumptions. The focus has been on using instrumental variables for identification of average treatment effects in a population of interest, or on the average effect for the sub-population that is treated. The conditions required to nonparametrically identify these parameters can be restrictive, however, and the derived identification results

fragile (Heckman [1990]). In particular, results in Chamberlain (1986), Manski (1990), and Angrist and Imbens (1991) require that there be some group in the sample for whom the probability of treatment is zero, at least in the limit.

The purpose of this paper is to show that even when there is no group available for whom the probability of treatment is zero, we can still identify an average treatment effect of interest, which we will call a *local average treatment effect* (LATE). This is the average effect of treatment for individuals whose treatment status is influenced by changing an exogenous regressor that satisfies an exclusion restriction. To obtain this result, we impose mild restrictions that are satisfied in a wide range of models and circumstances in economic research, including latent index models and evaluations based on *natural experiments* such as those by Angrist (1990) and Angrist and Krueger (1991). We do not make assumptions about the distribution of the response variables, nor do we assume that the treatment effect is constant. One interpretation of the result is that the incentives for participation are randomized, rather than the participation status itself. Combined with a restriction on the way incentives affect participation status this identifies the average treatment effect for those whose participation status can be changed by changing the incentives to participate.

In addition we analyze the probability limit of an instrumental variable estimator under these conditions. We show that it estimates a weighted average treatment effect, with the weights non-negative and adding up to one. If there is more than one instrument the probability limit potentially depends on the choice of the instrument. If the treatment effect

is identical for all individuals the choice of instrument is only important for efficiency.

The paper is organized as follows: In Section 2 we introduce the model. Section 3 contains the main identification result. In Section 4 we discuss estimation of local average treatment effects and the IV estimator. A number of examples are discussed in Section 5. The final section contains a summary of the main results and the conclusion.

2. The Model.

The framework we use is essentially similar to that outlined by Rubin (1974) and Heckman (1979) and described in our previous paper on identification of treatment effects (Angrist and Imbens [1991]). Let Y_0 be the response without the treatment or program for the typical individual.[2] Y_1 is the response with treatment. D is an indicator of treatment. We observe D and $Y = Y_D = D \cdot Y_1 + (1 - D) \cdot Y_0$ for a random sample of individuals. The individual treatment effect is $Y_1 - Y_0$ but Y_1 and Y_0 are never observed for the same person. Therefore we rely on comparisons between different individuals and compute average treatment effects. Two average treatment effects have received particular attention (see Heckman and Robb [1985]). The first is the average treatment effect for the entire population:

$$\alpha_0 = E[Y_1 - Y_0].$$

Outside of experimental situations, α_0 is not consistently estimated by taking the difference of the response for the individuals in the treatment and control group (or, equivalently, regressing Y on a constant and D) because such a difference has expectation equal to

[2] We will supress the index denoting the individual during most of the paper.

$$E[Y_1|D=1] - E[Y_0|D=0].$$

When individuals are not randomly assigned to the treatment and control groups, there is no reason to believe that $E[Y_0|D=0] = E[Y_0]$ and $E[Y_1|D=1] = E[Y_1]$.

A second average treatment effect of interest is the *selected average treatment effect*

$$\alpha_1 = E[Y_1 - Y_0|D=1].$$

This is the average effect of the treatment on the treated. Like α_0, α_1 is also difficult to identify because $E[Y_0|D=1]$ is not directly estimable.

3. IDENTIFICATION.

The solution to the identification problem dominating the evaluation of medical treatments is randomized assigment to treatment and control groups. This guarantees that $E[Y_i|D=0]$ is equal to $E[Y_i|D=1]$ for $i=0,1$. In the evaluation of social programs researchers have often relied on instrumental variables strategies to identify treatment effects. We define an instrumental variable Z to be a variable unrelated to the responses Y_0 and Y_1, and correlated with the participation probability. In order to formalize this, let $\{z_1, z_2, \ldots, z_K\}$ be the support of Z. Define for each z in the support of Z, a random variable D_z. D_z is equal to zero if an individual would not participate if he or she had the instrument Z equal to z, and it is equal to one if that individual would participate with $Z=z$. Clearly, we will not observe the entire set of indictators $\{D_{z_1}, D_{z_2}, \ldots, D_{z_K}\}$, but we assume that it exists in the same way that Y_0 and Y_1 are defined even if not observed.[3] We observe (Z, D, Y)

[3]The D_z notation was suggested to us by Gary Chamberlain.

for a random sample of individuals, where $D = D_Z = \sum_{k=1}^{K} I_{Z=z_k} \cdot D_{z_k}$, the participation indicicator associated with Z, and $Y = Y_D = D \cdot Y_1 + (1 - D) \cdot Y_0$, the variable given the participation status D.

The formal condition for instruments is:

Condition 1 (Existence of Instruments) *Let Z be any observed random variable such that Y_0, Y_1 and $D_{z_1}, D_{z_2}, \ldots, D_{z_K}$ are jointly independent of Z, and $P_z = E[D|Z = z] = E[D_z]$ is a non-trivial function of z.*

This condition guarantees that Z is a valid instrument in the sense that it does not directly affect the responses Y_0 and Y_1, and that it does affect the probability of participation in the program (i.e. that is is correlated with D).

It is important to note that Condition 1 by itself is not enough to identify *any* average treatment effect. The existence of a valid instrument implies that the endogeneity of treatment assigment can be dealt with, but it does not address the issue of treatment effect variation. The following example shows how severe the implication of treatment effect heterogeneity can be for the identification of an average treatment effect.

Example 1 Consider the following model:

$$Y_0 = \varepsilon$$

$$Y_1 = Y_0 + \eta$$

$$D_z = h(z, \nu)$$

and (ε, η, ν) are jointly independent of Z. If $Var(\eta) = 0$, a linear regression of Y on a constant and $P_z = E[D|Z = z]$ estimates the (constant) treatment effect. If the treatment effect η is not constant, this regression does not necessarily estimate anything of interest. Let Z be binary, and $\nu \in \{0, 1, 2\}$, with $Pr(\nu = 0) = 2/7$, $Pr(\nu = 1) = 4/7$ and $Pr(\nu = 2) = 1/7$. Further more, let $E[\eta|\nu = 0] = E[\eta|\nu = 1] = 1$, and $E[\eta|\nu = 2] = 2$. The conditional expectation of ε satisfied $E[\varepsilon|\nu = 0] = 1$, $E[\varepsilon|\nu = 1] = 5/2$, $E[\varepsilon|\nu = 2] = 0$. Finally, let the function $h(\cdot, \cdot)$ satisfy: $h(1, 0) = h(0, 2) = 1$, and $h(0, 0) = h(0, 1) = h(1, 1) = h(1, 2) = 0$. The participation probability P_z is $2/7$ for people with $Z = 0$ and $1/7$ for people with $Z = 1$. The average treatment effect α_0 is equal to $8/7$, and the selected average treatment effect α_1 equals $2/[Pr(Z = 0) + 1]$. However, the conditional expectation of Y given D and Z is equal to 2, no matter what the value of D or Z. The expected response is the same among participants as among non-participants, and it is the same among those with a high probability of participating ($Z = 0$) as among those with a low probability ($Z = 1$). Despite the fact that the treatment effect is positive for every individual, there is no comparison of the four average responses that could identify a meaningful treatment effect.

The reason there is no meaningful comparison is that shifting the instrument from 0 to 1 causes some people (those with $\nu = 0$) to shift from non-participation to participation, while others (with $\nu = 2$) shift from participation to non-participation. The treatment effect for those who shift to participation is cancelled out by the loss of those who shift from participation to non-participation for a given change in the instrument. □

One condition that could prevent cases like this is that the treatment effect is constant. Another condition is the existence of a value z of the instrument such that the probability of participation conditional on that value, P_z, is equal to zero. This type of condition is investigated in Heckman (1990) and Angrist and Imbens (1991). The next condition prevents shifts in participation status in opposite direction by limiting the variability in hypothetical participation decisions.

Condition 2 (Monotonicity) *For all z, w in the support of Z, such that $P_w \neq P_z$,*

$$(P_w - P_z) \cdot (D_w - D_z) \geq 0$$

If $P_w = P_z$ then $D_w = D_z$.

This condition ensures that the instrument affects the participation or selection decision in a monotone way. That is, if people are more likely to participate given $Z = w$ than given $Z = z$, then anyone who would participate given $Z = z$ must also participate given $Z = w$. Another way of phrasing Condition 2 is:

$$Pr(D_z - D_w \geq 0) = 1 \quad \text{or} \quad Pr(D_z - D_w \leq 0) = 1$$

Of course we can never verify this condition, because we observe people only with one value of Z, but in particular applications it might be a reasonable assumption.

An important class of models that satisfies Conditions 1 and 2 is the class of latent index models commonly used in econometric sample selection applications.

Example 2 *Consider the following model*

$$Y_0 = \mu + \varepsilon$$

$$Y_1 = Y_0 + \eta$$

$$D_z = I[z \cdot \gamma + \nu \geq 0]$$

Z is independent of (ν, ε, η).

The treatment effect η may be correlated with the stochastic component of both the response (ε) and the participation equation (ν), but all three must be independent of the instrument Z. Gronau (1974), Heckman (1978, 1990) and many others have used this type of model, often with variations, such as constant treatment effects (Var(η)=0), and with extra observable characteristics that affect both the participation and response. Heckman (1990) gives sufficient conditions for identification of the average treatment effect for the treated, α_1, from a random sample of (Y, D, Z). An important condition he requires is that the support of the instrument Z be unbounded. □

Our main result is the following:

Theorem 1 *If Conditions 1 and 2 hold, then we can identify the following average treatment effect:*

$$\alpha_{z,w} = E[Y_1 - Y_0 | D_z \neq D_w]$$

from the joint distribution of Y, D and Z, for all z and w in the support of Z such that $P_z \neq P_w$

Proof: See Appendix.

Theorem 1 implies that one can identify the average treatment effect for that part of the population that changes its participation behavior with the change in the instrument from $Z = z$ to $Z = w$. The relevance of the theorem depends on two issues. The first is whether the conditions are likely to be satisfied in practical evaluations of treatment effects. We give some examples below that suggest that there are cases where these conditions are satisfied. The other issue is whether the local average treatment effects that are identified here are of interest. Obviously if the treatment effect is identical for everybody, the local average treatment effect is equal to the average treatment effect for the population and for participants. In general, the less the treatment effect varies accross the population, the closer the local average treatment effect is to a population effect. Another point is that the local average treatment effect is the average treatment effect for the individuals whose behavior can be changed by changing the value of Z. The average treatment effect for this group might therefore be a good approximation for the treatment effect on those individuals who would be drawn into the program if it were to be made marginally more attractive. An appropriate analogy is to panel data techniques. In models with fixed effects, the data are only informative about the impact of binary regressors on individuals who change the value of the regressor over the period of observation. Here the treatment effect is identified only

for potential changers, those who can be induced to change participation status by a change in the instrument.

4. ESTIMATION.

The result in Theorem 1 suggests a simple procedure for estimating average treatment effects with a random sample of Y, D and Z. when Z is binary with values z and w. In this case the only treatment effect identified is α_{zw}. From the proof of Theorem 1 we have:

$$\alpha_{zw} = E[Y_1 - Y_0 | D_z \neq D_w] = \{E[Y|Z = z] - E[Y|Z = w]\}/\{P_z - P_w\}$$

which can be consistently estimated by replacing expectations with sample averages. The resulting estimator is an application of Wald's (1940) method of fitting straight lines and is the same as an instrumental variables estimator using binary Z and a constant as instruments (Durbin [1954]).

If Z has a discrete distribution with points of support $z_0, z_1, \ldots, z_K$ we can estimate $(K+1) \times K/2$ average treatment effects, only K of which are linearly independent. They are related as follows:

$$\alpha_{z_m,z_k} = \frac{P_{z_l} - P_{z_k}}{P_{z_m} - P_{z_k}} \alpha_{z_l,z_k} + \frac{P_{z_m} - P_{z_l}}{P_{z_m} - P_{z_k}} \alpha_{z_m,z_l}$$

Let the points of support be ordered in such a way that $l < m$ implies $P_l < P_m$. Then the average treatment effect that refers to the largest group is $\alpha_{z_K z_0}$.[4] This can be expressed in

[4] α_{zw} is the average treatment effect for a group that makes up a proportion equal to $|P_z - P_w|$ of the total population. The size of this proportion is maximized by taking the treatment effect $\alpha_{z_K z_0}$.

terms of a complete set of K linearly independent treatment effects based on first differences as:

$$\alpha_{z_K z_0} = \Big[\sum_{k=1}^{K}(P_{z_k} - P_{z_{k-1}}) \cdot \alpha_{z_k z_{k-1}}\Big] \Big/ \Big[P_{z_K} - P_{z_0}\Big]$$

A sequence of K estimated local average treatment effects can be used to assess the variability of treatment effects. When the effects appear stable they can be combined into a single efficient estimate.

Angrist (1991) discusses a simple framework for imposing and testing the constant treatment effect hypothesis on a complete set of linearly independent Wald estimates. The efficient linear combination of any set of K linearly independent Wald estimates is computed by using K indicators for each value of Z, $\delta_k = I[Z = z_k]$ as instruments for P along with a constant in conventional Two-Stage Least Squares estimation of the equation:

$$(1) \quad Y = E[Y_0] + \alpha_\lambda \cdot P + \varepsilon = E[Y_0] + \alpha_\lambda \cdot P + \{[(Y_1 - Y_0) - \alpha_\lambda] \cdot P + Y_0 - E[Y_0]$$

This is the same as using P_z as an instrument for P in (1). Finally, the same estimate of α_λ can be computed by Generalized Least Squares estimation of the grouped equation

$$(2) \quad Y_k = E[Y_0] + \alpha_\lambda \cdot P_{z_k} + \bar{\nu}_k$$

where $\bar{\nu}_k$ is the average of the compound error term given $Z = z_k$, and $Y_k = E[Y|Z = z_k]$.

A Wald test for the equality of a full set of linearly independent Wald estimates is the Chi-square goodness of fit statistic for (2). Alternatively, the same statistic can be computed

as the standard instrument–error orthogonality test statistic (see, e.g. Newey [1985]) using the set of instruments, δ_k (Angrist 1991, Proposition 2).

In empirical applications, local average treatment effects will often differ. Even in this case the parameter estimated by TSLS using a constant and P_z or a full set of dummies as instruments applied to equation (1) is still of interest. It can be shown that the probability limit of the TSLS or IV estimator of α_λ is

$$(3) \quad \alpha_\lambda = \frac{E[Y \cdot (P_z - E[P_z])]}{E[P_z \cdot (P_z - E[P_z])]} = \sum_{k=1}^{K} \lambda_k \cdot \alpha_{z_k z_{k-1}}$$

with

$$\lambda_k = (P_{z_k} - P_{z_{k-1}}) \cdot \frac{\sum_{l=k}^{K} \pi_l \cdot (P_{z_l} - Q)}{\sum_{l=0}^{K} \pi_l \cdot P_{z_l} \cdot (P_{z_k} - Q)}$$

where $\pi_k = Pr(Z = z_k)$ and $Q = Pr(D = 1) = \sum_{k=0}^{K} \pi_k P_{z_k}$. This probability limit α_λ is a weighted average of the K average treatment effects $\alpha_{z_k z_{k-1}}$ with the weights λ_k non–negative and adding up to one.[5]

If $P_{z_0} = 0$, then the conditions for identification of the average treatment effect on the treated are satisfied. We have:

$$\alpha_1 = \sum_{k=1}^{K} \frac{\pi_k P_{z_k}}{Q} \cdot \alpha_{z_k z_0}$$

If Z is continuous one could parametrize the treatment effect, for instance by dividing the support of Z in intervals and estimating an average treatment effect for every interval.[6]

[5]This follows from the ordering of the points of support of Z in such a way that $P_{z_k} \geq P_{z_{k-1}}$ for all k.

[6]Condition 1 now requires that Z is independent of Y_0, Y_1 and the infinite set of D_z for all z in the support of Z.

Alternatively one can regress Y on a constant and P_Z, equivalent to using P_z and a constant as instruments for P. This is feasible if P_z can be estimated. In investigating the probability limit of the estimator of the coefficien on P_Z in such a regression we will assume that Z is a one–dimensional random variable with density $f(z)$. In addition we assume that P_z is a differentiable function of z with positive derivative. These assumptions might seem more restrictive than they really are. If Z is not one–dimensional, or P_z is not non–decreasing, one can transform the instrument from Z to $\tilde{Z} = P_Z$, with participation probability $\tilde{P}_z = E[D|\tilde{Z} = z]$. $\tilde{Z}$ satisfies Conditions 1 and 2, is one–dimensional and $\tilde{P}_z$ is non–decreasing because $\tilde{P}_z = E[D|\tilde{Z} = z] = E[D|P_Z = z] = z$. Also, the regression of Y on $\tilde{P}_Z$ and the regression of Y on P_Z give identical results because $P_z = E[D|Z = z] = E[D|\tilde{Z} = P_z] = \tilde{P}_{P_z}$. Define:

$$\alpha_z = \lim_{w \uparrow z} \alpha_{zw} = E[Y_1 - Y_0 | \lim_{w \uparrow z} D_z - D_w \neq 0]$$

We assume that this limit is well defined. α_z is the expected treatment effect for someone who will change from participation to non–participation if the value of the instrument is lowered from its original value of z, however small the decrease. An example should help to clarify the nature of α_z. In the latent index model of example 2,

$$\alpha_z = E[Y_1 - Y_0 | \lim_{w \uparrow z} D_z - D_w \neq 0]$$

$$= E[Y_1 - Y_0 | \lim_{w \uparrow z} I[z \cdot \gamma + \nu \geq 0] - I[w \cdot \gamma + \nu \geq 0] \neq 0]$$

$$= E[Y_1 - Y_0 | \lim_{w \uparrow z} -w \cdot \gamma > \nu \geq -z \cdot \gamma]$$

$$= E[Y_1 - Y_0|\nu = -z \cdot \gamma]$$

The main result in this section is the following

Theorem 2 *The probability limit of the estimator for the coefficient on P_z in the regression of Y on a constant and P_z, or the IV estimator using P_z as an instrument, is equal to:*

$$\alpha_\lambda = \frac{E[Y \cdot (P_z - Q)]}{E[P_z \cdot (P_z - Q)]} = \int_{-\infty}^{\infty} \alpha_z \cdot \lambda(z) dz$$

with

$$\lambda(z) = \left[\frac{\partial P_z}{\partial z} \cdot \int_z^{\infty} (P_u - Q) \cdot f(u) du\right] \Big/ \left[\int_{-\infty}^{\infty} P_u \cdot (P_u - Q) \cdot f(u) du\right]$$

The weight function is non-negative and integrates to one. The weights are proportional to the derivative of P_z and to $\int_z^{\infty}(P_u - Q) \cdot f(u)du$.

Proof: See Appendix.

An important implication of the above result is that different instruments may lead to different estimates α_λ if the treatment effect is not constant. The reason is not necessarily that the instruments are not valid but may be that the weight functions associated with the different estimators are different. This result might explain some of the variation found in estimated treatment effects, such as union wage effects (cf Lewis [1986]). The following example shows how the variation in treatment effects might affect different IV estimators.

Example 3 Let $Y_0 = 0$, $Y_1 = \eta$. Z_1 and Z_2 are both potential instruments, and are independent binary random variables. Let D be equal to $h(Z_1, Z_2, \nu)$, with

$$h(0,0,0) = h(0,0,1) = h(0,1,0) = h(1,0,1) = 0$$

$$h(0,1,1) = h(1,0,0) = h(1,1,0) = h(1,1,1) = 1$$

The treatment effect η has conditional expectation $E[\eta|\nu = 0] = \eta_0$ and $E[\eta|\nu = 1] = \eta_1$.

This model satisfies the two conditions in two ways. From the point of view of the researcher with instrument Z_1, $Z = Z_1$ and $D_z = h(z, Z_2, \nu)$. For this researcher Condition 1 requires that Y_0, Y_1 and $D_z = h(z, Z_2, \nu)$ are independent of $Z = Z_1$, and Condition 2 requires that $\{E[h(1, Z_2, \nu)] - E[h(0, Z_2, \nu)]\} \cdot [h(1, Z_2, \nu) - h(0, Z_2, \nu)] \geq 0$ for all Z_2 and ν. This researcher will estimate the local average treatment effect (which in the binary instrument case is equal to the treatment effect estimated by the IV estimator) to be equal to $(E[Y|Z_1 = 1] - E[Y|Z_1 = 0])/(E[D|Z_1 = 1] - E[D|Z_1 = 0]) = \eta_0 \cdot (1 - Pr(\nu = 1)) + \eta_1 \cdot Pr(\nu = 1)$. From the point of view of the second researcher with instrument Z_2, $Z = Z_2$ and $D_z = h(Z_1, z, \nu)$. Condition 1 and 2 change accordingly. He or she will estimate the local average treatment effect to be $(E[Y|Z_2 = 1] - E[Y|Z_2 = 0])/(E[D|Z_2 = 1] - E[D|Z_2 = 0]) = \eta_1$.

The reason for the difference in estimated treatment effects is that the instruments lead to different weights for the treatment effects η_0 and η_1. □

If there are additional variables X which affect both the response and probability of participation there are a number of possibilities. The first is that X is uncorrelated with Z. In that case X is just part of the stochastic component that is common to the participation and response equation. One can leave it out of the analysis completely, since it does not

affect the estimates for the treatment effects. It might however reduce the variance of such estimates by increasing the R^2 of the response regression. The second case is that where X is correlated with Z as well as Y. One could modify Condition 1 to state that Y_0, Y_1 and D_z are independent of Z, conditional on X. The entire analysis could then be done conditional on X, either non-parametrically, or by assuming a particular parametric form for the dependence on X.

5. EXAMPLES.

In this section we will give a number of examples in which conditions 1 and 2 are likely to be satisfied. The examples exploit the manner in which a particular program or treatment is implemented to create instruments that clearly are exogenous. Evaluations of this type are sometimes reffered to as *natural experiments*, in contrast with the identification achieved in clinical trials where individuals are directly randomized in or out of treatment and control groups. The difference between natural experiments and clinical trials is not fundamental, rather it is related to the stage at which randomization occurs.

Example 4 (Draft Lottery) Angrist (1990) uses the Vietnam-era draft lottery to identify the earnings effect of veteran status on earnings. The instrument is the draft lottery number, randomly assigned to date of birth. Because they were randomly assigned, lottery numbers do not directly affect the response variable (be it earnings or employment status). However, lottery numbers were used to determine priority for conscription (veteran status itself was never randomly assigned). Therefore the average probability of serving in the military falls

with the lottery number. Condition 2 essentially requires that someone who would serve in the military with lottery number k, would also serve in the military with lottery number $l < k$.

An example of the average effect of veteran status that is identified is that for people who would serve if they had a low lottery number, but not with a high lottery number. One cannot say anything about the effect of veteran status on people who would serve regardless of their lottery numbers (i.e. true volunteers), nor can one say anything about people who would not serve regardless of their lottery number (draft ineligible men, draft evasions). The average treatment effect is identified for the people affected by the draft. □

Example 5 (Compulsory Schooling Age) Angrist and Krueger (1991) investigate the effect of schooling on earnings using the variation in compulsory schooling created by the variation in birth dates. Birth dates themselves probably do not affect earnings in cohorts born in a period that is not too long. Birth dates do affect the level of schooling achieved because people born on different dates, who are allowed to drop out of school on their birthday, are confronted with (slightly) different compulsory schooling levels. Condition 2 here requires that someone who would stay in school when not compelled, would also stay in school when constrained to do so by accident of birth. The local average treatment effect identified here is that of people who are (potentially) affected by the compulsory schooling laws. □

The following example illustrates that condition 2 is not trivially satisfied in practice.[7]

Example 6 (Administrative Screening) Suppose applicants for a social program are screened by a number of officials. Different officials have different admission rates, even if the stated admission criteria are identical. Since the identity of the official is clearly immaterial to the response, it satisfies Condition 1. However, Condition 2 requires that if we have two officials, one who accepts applicants with probability p_0, and a second one who accepts people with probability $p_1 > p_0$, the second official must accept any applicant who would have been accepted by the first official, or $D_1 \geq D_0$. This is unlikely to hold in practice if admission is decided on a number of criteria. Therefore we cannot use Theorem 1 to identify a local average treatment effect despite the presence of a valid instrument that does not affect response, but does affect the participation probability. IV estimation might in this case reflect different screening procedures rather than any real treatment effect. □

The final example deals with an important epidemiological problem, randomization of "intention to treat". It is taken from Robins (1989).

Example 7 (Randomization of Intention to Treat) Let Z be equal to 0 or 1 if a particular individual is assigned to the control or treatment group. The actual treatment indicator D may differ from Z because some individuals may not comply with the assigment. For this case, Robins (1989) lays out different assumptions that identify or give bounds on the average treatment effects for the entire population or for the treated. Condition 2 here

[7]This example was suggested to us by Geert Ridder.

requires that people who were assigned treatment but did not receive it, would also not receive treatment if they were not assigned to it.[8] In addition, people who received treatment despite being assigned to the control group, would also receive it if they were assigned to the treatment group. This seems likely to hold in practice if non-compliance is the result of concious behavior by the patients. Formally, Condition 2 requires that $D_1 \geq D_0$.

The treatment effect identified here is the average treatment effect for those who always comply with their assigment. □

6. Conclusion.

In this paper we have given sufficient conditions for the identification of a class of average treatment effects. As with previous exclusion restrictions, these conditions require the existence of an instrument that shifts the participation probability, but does not directly affect the response. In the case of heterogeneous treatment effects, exclusion restrictions may not be enough to identify the average treatment effect on the treated, but they are enough to identify the average treatment effect for a subset of the participants. The most important part of an empirical research agenda remains uncovering the potential instruments in the first place. Elsewhere (Angrist and Imbens [1991]) we have argued that this is often a matter of careful investigation of the implementation of a particular program.

The purpose of ths paper is to show applied researchers that given a valid instrument, technical considerations arising from treatment effect heterogeneity need not inhibit in-

[8] Assuming that the marginal probability of receiving treatment is higher if someone was actually assigned to it than if he or she was assigned to the control group, or $P_0 < P_1$.

ference. Moreover, we show that a conventional 2SLS estimator consistently estimates a weighted average of local average treatment effects in the class of models identified under our main theoretical result.

Appendix

Proof of Theorem 1

From the joint distribution of Y, D and Z we can identify $Y_z = E[Y|Z = z]$, $Y_w = E[Y|Z = w]$, $P_z = E[D|Z = z]$ and $P_w = E[D|Z = w]$. We will show that α_{zw} is equal to $(Y_z - Y_w)/(P_z - P_w)$.

Consider the numerator $Y_z - Y_w$. It is equal to

$$E[Y|Z = z] - E[Y|Z = w]$$

$$= E[D_Z \cdot Y_1 + (1 - D_Z) \cdot Y_0|Z = z] - E[D_Z \cdot Y_1 + (1 - D_Z) \cdot Y_0|Z = w]$$

$$= E[D_z \cdot Y_1 + (1 - D_z) \cdot Y_0|Z = z] - E[D_w \cdot Y_1 + (1 - D_w) \cdot Y_0|Z = w]$$

$$= E[D_z \cdot Y_1 + (1 - D_z) \cdot Y_0] - E[D_w \cdot Y_1 + (1 - D_w) \cdot Y_0]$$

$$= E[D_z \cdot Y_1 + (1 - D_z) \cdot Y_0 - D_w \cdot Y_1 + (1 - D_w) \cdot Y_0]$$

$$= E[(D_z - D_w) \cdot (Y_1 - Y_0)]$$

$$(4) \quad = Pr(D_z - D_w = 1) \cdot E[Y_1 - Y_0|D_z - D_w = 1]$$

$$- Pr(D_z - D_w = -1) \cdot E[Y_1 - Y_0|D_z - D_w = -1]$$

Sofar we have not used Condition 2. We did use the independence implied by Condition 1. If $P_z > P_w$, Condition 2 implies that $D_z - D_w \geq 0$, and therefore the second term of the last equation is equal to zero. Then:

$$E[Y|Z=z] - E[Y|Z=w] = Pr(D_z - D_w = 1) \cdot E[Y_1 - Y_0|D_z - D_w = 1]$$

$$= (P_z - P_w) \cdot E[Y_1 - Y_0|D_z \neq D_w]$$

Therefore $(Y_z - Y_w)/(P_z - P_w)$ is equal to $E[Y_1 - Y_0|D_z \neq D_w]$. If $P_z < P_w$ Condition 2 implies that $D_z \leq D_w$, and the first term in (4) is equal to zero. The same procedure as above then leads to the same result. $\mathcal{QED}$.

Proof of Theorem 2

The estimator for the coefficient on P_z in the regression of Y on a constant and P_z, with a random sample of (Y, P_z) of size N, is equal to:

$$\frac{\sum_{n=1}^{N} Y_n \cdot (P_{z_n} - \bar{P}_z)}{\sum_{n=1}^{N} P_{z_n} \cdot (P_{z_n} - \bar{P}_z)}$$

where $\bar{P}_z = \sum_{n=1}^{N} P_{z_n}/N$. The probability limit of this estimator is equal to

$$\frac{E[Y \cdot (P_z - Q)]}{E[P_z \cdot (P_z - Q)]}$$

where $Q = E[P_z]$. We have to show that

$$\frac{E[Y \cdot (P_z - Q)]}{E[P_z \cdot (P_z - Q)]} = \frac{\int_{-\infty}^{\infty} \int_{z}^{\infty} \alpha_z \frac{\partial P_z}{\partial z} \cdot (P_u - Q) \cdot f(u) du dz}{\int_{-\infty}^{\infty} P_z \cdot (P_z - Q) \cdot f(z) dz}$$

The numerators of both sides are clearly identical, so the issue is whether

$$(5) \quad E[Y \cdot (P_z - Q)] = \int_{-\infty}^{\infty} \int_{z}^{\infty} \alpha_z \frac{\partial P_z}{\partial z} \cdot (P_u - Q) \cdot f(u) du$$

The first step is to analyze the conditional expectation of (5) given $Z = z$:

$$(6) \quad E[Y|Z = z] = E[Y_0 + P \cdot (Y_1 - Y_0)|Z = z] = E[Y_0] + P_z \cdot E[Y_1 - Y_0|Z = z, D = 1]$$

$$= E[Y_0] + P_z \cdot E[Y_1 - Y_0|Z = z, D_z = 1] = E[Y_0] + P_z \cdot E[Y_1 - Y_0|D_z = 1]$$

The treatment effect $E[Y_1 - Y_0|D_z = 1]$ is related to the local average treatment effects by the following equation:

$$(7) \quad E[Y_1 - Y_0|D_z = 1] = Pr(D_w \neq D_z|D_z = 1) \cdot E[Y_1 - Y_0|D_w \neq D_z]$$

$$+Pr(D_w = 1|D_z = 1) \cdot E[Y_1 - Y_0|D_w = 1]$$

$$= \frac{P_z - P_w}{P_z} \cdot \alpha_{zw} + \frac{P_w}{P_z} \cdot E[Y_1 - Y_0|D_w = 1]$$

for all values of $w < z$ in the support of Z. We have used here the fact that $w < z$ implies $P_z > P_w$ and therefore $D_w = 1$ implies $D_z = 1$. We can use (7) to calculate the derivative of (6) with respect to z, It is equal to:

$$\lim_{w \uparrow z} \frac{E[Y|Z = z] - E[Y|Z = w]}{z - w} = \lim_{w \uparrow z} \frac{P_z - P_w}{z - w} \cdot \alpha_{zw} = \frac{\partial P_z}{\partial z} \cdot \alpha_z$$

This implies that

$$E[Y|Z = z] = c + \int_{-\infty}^{z} \frac{\partial P_u}{\partial u} \cdot \alpha_u du$$

for some constant c. Hence:

$$E[Y \cdot (P_z - Q)] = \int_{-\infty}^{\infty} E[Y \cdot (P_z - Q)|Z = z] \cdot f(z)dz$$

$$= \int_{-\infty}^{\infty} \Big(c + \int_{-\infty}^{z} \frac{\partial P_u}{\partial u} \cdot \alpha_u du\Big) \cdot (P_z - Q) \cdot f(z)dz$$

$$= \int_{-\infty}^{\infty} \int_{-\infty}^{z} \frac{\partial P_u}{\partial u} \cdot \alpha_u \cdot (P_z - Q) \cdot f(z)dudz$$

$$= \int_{-\infty}^{\infty} \int_{z}^{\infty} \alpha_z \frac{\partial P_z}{\partial z} \cdot (P_u - Q) \cdot f(u)dudz$$

This proves the first part of the Theorem. To see that the weights are non–negative, note that $\frac{\partial P_z}{\partial z}$ is positive by assumption. Also,

$$\int_{z}^{\infty} (P_u - Q) \cdot f(u)du = -\int_{-\infty}^{z} (P_u - Q) \cdot f(u)du$$

If $P_z > Q$, then the left hand side is positive because $P_u > Q$ for all $u \geq z$. If $P_z < Q$, then the right hand side is positive because $P_u < Q$ for all $u \leq z$. So, in both cases the left hand side is greater than or equal to zero.

The final part is to show that the weight function intergrates out to one:

$$\int_{-\infty}^{\infty} \lambda(z)dz$$

$$= \Big[\int_{-\infty}^{\infty} \frac{\partial P_z}{\partial z} \cdot \int_{z}^{\infty} (P_u - Q) \cdot f(u)dudz\Big] \Big/ \Big[\int_{-\infty}^{\infty} P_u \cdot (P_u - Q) \cdot f(u)du\Big]$$

$$= \Big[\int_{-\infty}^{\infty} \int_{-\infty}^{u} \frac{\partial P_z}{\partial z} \cdot (P_u - Q) \cdot f(u)dzdu\Big] \Big/ \Big[\int_{-\infty}^{\infty} P_u \cdot (P_u - Q) \cdot f(u)du\Big]$$

$$= \Big[\int_{-\infty}^{\infty} P_u \cdot (P_u - Q) \cdot f(u)du\Big] \Big/ \Big[\int_{-\infty}^{\infty} P_u \cdot (P_u - Q) \cdot f(u)du\Big] = 1$$

$\mathcal{QED}$.

REFERENCES

ANGRIST, J.D., (1990). "Lifetime Earnings and the Vietnam Era Draft Lottery: Evidence from Social Security Administrative Records," *American Economic Review*, June 1990.

—, (1991), "Grouped-Data Estimation and Testing in Simple Labor-Supply Models", *Journal of Econometrics*, 47, 243–266.

—, (1991), "Linear Instrumental variables Estimation of Average Treatment Effects in Nonlinear Models," Harvard Institute for Economic Research Discussion Paper No. 1542, March 1991.

—, AND A. KRUEGER, (1991), "Does compulsory school attendance affect schooling and earnings", Harvard Institute of Economic Research Discussion paper No. 1475, March.

—, AND G. IMBENS, (1991), "Sources of Identifying Information in Evaluation Models", Department of Economics, Harvard University.

CHAMBERLAIN, G., (1986), "Asymptotic efficiency in semi-parametric models with censoring", *Journal of Econometrics*, 32; 189–218.

DURBIN, J., (1954), "Errors in Variables", *Review of the International Statistical Institute*, 22: 23–32.

GRONAU, (1974), "Wage Comparisons – A Selectivity Bias", *Journal of Political Economy*, 82: 1119–43.

HECKMAN, J. J., (1979), "Sample Selection Bias as a Specification Error", *Econometrica*, 47(1);153–161.

—, (1990a), "Varieties of Selection Bias," *American Economic Review*, 80(2) : 313-318.

—, AND J. HOTZ (1989), "Choosing Among Alternative Nonexperimental Methods for Estimating the Impact of Social programs: The Case of Manpower Training," *Journal of the American Statistical Association*, 84 (December 1989), 862-874.

—, AND R. ROBB, (1985), "Alternative Methods for Evaluating the Impact of Interventions," in J. Heckman and B. Singer, eds., Longitudinal Analysis of Labor Market Data, New York: Cambridge University Press.

LALONDE, R. J., (1986), " Evaluating the Econometric Evaluations of Training Programs Using Experimental Data," *American Economic Review* 76(4): 602-620.

LEWIS, H. G., (1986), *Union Relative Wage Effects*, Chicago, University of Chicago Press.

MANSKI, C. F., (1990b), "The Selection Problem," University of Wisconsin, Social Systems Research Institute Working paper 9012, September.

NEWEY, W., (1985), "Generalized Method Of Moments Estimation and Testing", *Journal of Econometrics*, 29: 229–256.

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

ROYALL, M., (1991), "Ethics and Statistics in Randomized Clinical Trials", *Statistical Science*, 6(1): 52–88.

RUBIN, D. (1974), "Estimating Causal Effects of Treatments in Randomized and Non-randomized Studies," *Journal of Educational Psychology*, 66(5): 688- 701.

WALD, A., (1940), "The Fitting of Straight Lines if Both Variables are Subject to Error", *Annals of Mathematical Statistics*, 11: 284–300.