NBER TECHNICAL WORKING PAPERS SERIES

SOURCES OF IDENTIFYING INFORMATION IN EVALUATION MODELS

Joshua D. Angrist

Guido W. Imbens

Technical Working Paper No. 117

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

We are grateful to Charles Manski, Robert Moffit and Geert Ridder for helpful comments, and especially to Gary Chamberlain for many illuminating discussions. Part of the paper was written while the second author was visiting the Center for Economic Research at Tilburg University. 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 #117
December 1991

SOURCES OF IDENTIFYING INFORMATION IN EVALUATION MODELS

ABSTRACT

The average effect of social programs on outcomes such as earnings is a parameter of primary interest in econometric evaluations studies. New results on using exclusion restrictions to identify and estimate average treatment effects are presented. Identification is achieved given a minimum of parametric assumptions, initially without reference to a latent index framework. Most econometric analyses of evaluation models motivate identifying assumptions using models of individual behavior. Our technical conditions do not fit easily into a conventional discrete choice framework, rather they fit into a framework where the source of identifying information is institutional knowledge regarding program administration. This framework also suggests an attractive experimental design for research using human subjects, in which eligible participants need not be denied treatment. We present a simple instrumental variables estimator for the average effect of treatment on program participants, and show that the estimator attains Chamberlain's semi-parametric efficiency bound. The bias of estimators that satisfy only exclusion restrictions is also considered.

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. Do programs that subsidize education and training improve the labor market outcomes of program participants? Evaluation questions of this type are of great concern to government policy makers, private employers, and academic researchers. In any field where scientific research has policy implications, evaluation methodology is also of considerable importance. Discussions of evaluation methodology are discussions of the nature and credibility of scientific evidence. In medical research, for example, government regulations establish standards and procedures that researchers must follow for their results to be considered credible evidence for the efficacy and safety of new drugs. Standards here are quite clear: research guidelines for a new drug application clearly favor, but do not require, the randomized assigment of treatment and concurrent data collection on control groups (Center for Drug Evaluation and Research 1988, pp. 22, 56)

Social policy is arguably as important for human welfare as public health, yet no mutually agreed standard of evidence exists for establishing the effectiveness of social programs. On the one hand, critical research on econometric evaluation methodology by Lalonde (1986) and others has led to renewed interest in classical experimentation as a tool for social policy evaluations. Manski and Garfinkel (1991) note that the recent Job Partnership Training Act (JPTA) even mandates a particular sort of treatment–control evaluation design in which applicants for training are randomly denied treatment. On the other hand, Manski and Garfinkel (1991) and Heckman and Hotz (1989) argue persuasively that experiments can never be a complete substitute for evaluations using observational data. Disagreements over

evaluation methodology notwithstanding, research directed towards adapting experimental designs for social policy analysis and allowing for fewer assumptions in observational analyses is likely to remain important. This paper contributes to both the experimental and observational components of the evaluation research agenda by presenting new results on using exclusion restrictions to identify and estimate average treatment effects.

Our findings are related to results in a number of recent papers on theoretical identification in evaluation models. Like Chamberlain (1986), Heckman (1990a) and Heckman and Honoré (1990), we are concerned with identification given a minimum number of parametric assumptions. But, as in Manski (1990), we avoid the additive latent index framework commonly invoked in econometric evaluations. Much of the previous work on identification presents some very general findings regarding the identification of distributions, but devotes relatively little attention to converting theoretical identification into empirically feasible estimators. In contrast, the formulation in this paper focuses on conditional means, and is immediately useful to applied researchers because it provides necessary and sufficient conditions for linear instrumental variables techniques to consistently estimate the average effect of treatment. In this, our approach is related to Angrist's (1991) use of instrumental variables to estimate treatment effects in nonlinear models, although here the identification conditions are not motivated by functional form restrictions.

We also show how to interpret the identifying assumptions as outlining a particular type of experimental design useful for research involving human subjects. Like Heckman (1990b),

we view social experiments as a source of identifying information, rather than as a replacement for economic modelling, and think that experiments should be designed with this in mind. An experimental design interpretation of Instrumental Variables identification conditions is important because the resulting design may be ethically more attractive than the conventional approach to randomization wherein eligible program applicants are randomly excluded from treatment. For example, some physicians have argued that randomization is incompatible with the Personal Care Principle in medical ethics, which requires doctors to put the welfare of their patients above the potential social gains from research (Royall [1991]). JPTA program administrators are also reluctant to deny training to applicants randomized into a control group (Hotz [1991]).

Our framework for experimental design essentially consists of first choosing an eligible population or evaluation site, either by randomized manipulation, or on the basis of ignorable (as defined by Rosenbaum and Rubin [1983]) covariates. Any eligible participant is then allowed to participate in the program if he or she likes. This approach may also identify parameters which are more likely to be useful for forecasting the impact of future programs.[5]

As a related by-product, our approach to inference also provides some insight regarding the problem of non–compliance in clinical trials, recently analyzed by Efron and Feldman (1991) and Robins (1989). Randomization of intention–to–treat, but not actual treatment,

[5]Harris (1985) and Moffit (1991b) also discuss randomization of sites versus randomization of individuals. However, a key distinction is that within sites these authors argue for saturation of treatment within sites while we do not. Different average treatment effects are therefore identified in the two types of site-randomization designs.

is one way to generate exclusion restrictions that will be sufficient to identify an average treatment effect. Not surprisingly, the estimator that uses these exclusion restrictions is a form of instrumental variables.

The paper is organized as follows. Section 2 formally defines the average effect of treatment on program participants and presents the main theoretical results. Necessary and sufficient conditions are given for a data generating process to identify an average treatment effect under exclusion restrictions. These results are also compared to previous results on the identification of treatment effects. Section 3 outlines the instrumental variables interpretation of identifying information and discusses the type of experimental design or data generating processes that satisfy the identifying conditions. Some results on the efficient use of exclusion restrictions in estimating average treatment effects are also discussed. In section 4 we discuss what can be learned about treatment effects if the average treatment effect of interest is not identified. It is shown that one might still be able to derive bounds for the average treatment effect. Section 5 offers a summary and some concluding thoughts on the nature of identifying information in models for the evaluation of social programs. An important distinction, and an underlying theme of the paper, is the difference between identifying information derived from models of program participants' behavior and from information about program eligibility rules. We argue that the latter is more likely to provide a convincing empirical identification strategy.

2.1 IDENTIFICATION. Our framework is essentially similar to that advanced by Rubin (1977),

Heckman (1990), and others. Let Y_0 be the response variable for an individual if he or she does not participate in the program. We assume that Y_0 is well defined even if the individual is actually participating in the program. Similarly, Y_1 is the value of the response variable if the individual does participate in the program, and $Y_1 - Y_0$ is the treatment effect that we are interested in. We never observe both Y_0 and Y_1; all inferences about these differences are indirect and in terms of expectations. Let $f_0(y)$ and $f_1(y)$ denote the probability density functions of Y_0 and Y_1 respectively. P denotes an indicator for program participation, equal to one if an individual participates in the program, and equal to zero otherwise.

The average treatment effect can be defined in a number of ways (See, e.g., Heckman and Robb [1985], and Heckman [1990]). First, there is the expectation of $Y_1 - Y_0$ in the population:

$$(1) \quad \bar{\alpha} = E[Y_1 - Y_0] = \int y[f_1(y) - f_0(y)]dy$$

This is the expected treatment effect if we take an individual randomly from the population and look at the difference between his response as a participant and nonparticipant. A second average treatment effect is defined by taking the expectation conditional on participation:

$$(2) \quad \alpha = E[Y_1 - Y_0|P = 1] = \int y[f_1(y|P = 1) - f_0(y|P = 1)]dy$$

This measures how much a participant gains from the program. Whether the focus is on the average treatment effect (ATE), $\bar{\alpha}$, or on the selected average treatment effect (SATE), α, depends on the particular application. We are usually interested in forecasting the effects

of a program when it is extended to a larger part of society. If the program or treatment will potentially be used by all members of the population, $\tilde{\alpha}$ is appropriate. If the program will eventually be used by a population with characteristics similar to the population in the evaluation design, α is the relevant average treatment effect. The latter is probably more realistic in economic applications. We will therefore concentrate on identification of α, rather than $\tilde{\alpha}$.

The problem of estimating average treatment effects in our framework is one of sample selection exactly the same as that considered by Gronau (1974), Heckman (1979) and Manski (1990). We observe $Y = P \cdot Y_1 + (1 - P) \cdot Y_0$ and P. From this two conditional response distributions are identified:

$$f_1(y|P = 1) \quad \text{and} \quad f_0(y|P = 0),$$

along with the probability of participation, $q = Pr(P = 1)$. These distributions do not allow us to calculate $\tilde{\alpha}$ or α, for which we need to know the counterfactual expectation $E[Y|P = 1]$. The difference between the mean of Y_1 for those who participate and Y_0 for those who do not participate can be written as

$$E[Y_1|P = 1] - E[Y_0|P = 0]$$

$$= E[Y_1 - Y_0|P = 1] + E[Y_0|P = 1] - E[Y_0|P = 0] = \alpha + \beta$$

The average difference in outcomes between program participants and nonparticipants generally confounds the treatment effect α and the selection effect β. The exception is when

$f_i(y|P=1)$ is equal to $f_i(y|P=0)$ for $i=0,1$ and all y, in which case selection is sometimes said to be ignorable. This implies that the two response distributions (with and without participation) do not depend on the decision to participate. If this is not the case then selection is non–ignorable and it is clear that we need more information, or restrictions on $f_0(\cdot)$, to separate α and β. Below, we briefly review some identifying assumptions.

The first approach assumes that the selection problem can be solved simply by conditioning on the right covariates.

Condition 1 *There is an observable covariate X such that*

$$E[Y_i|P=1,X=x]=E[Y_i|P=0,X=x]$$

In this case we can condition on X to remove the selection effect if we observe (Y,P,X):

$$\alpha=\int E[Y_1|P=1,X=x]-E[Y_0|P=1,X=x]\cdot g(x|P=1)dx$$

$$=\int E[Y_1|P=1,X=x]-E[Y_0|P=0,X=x]\cdot g(x|P=1)dx$$

This is in terms of expectations and distributions that can usually be estimated. The selection effect is equal to:

$$\beta=\int E[Y_0|P=1,X=x]\cdot g(x|P=1)-E[Y_0|P=0,X=x]\cdot g(x|P=0)]dx$$

$$=\int E[Y|P=0,X=x]\cdot[g(x|P=1)-g(x|P=0)]dx$$

which can also be estimated. If $g(x|P=1) = g(x|P=0)$ for all x, implying that $Pr(P=1|x)$ does not depend on x, selection is ignorable after all and the selection effect is zero.

Conditioning on covariates corresponds to identification by adequately controlling for all factors related to both outcomes and treatment. References for this approach include Rubin (1977) and, in a regression framework, Barnow, Cain, and Goldberger (1981). A generalized control function methodology is outlined by Heckman and Robb (1985). An alternative approach to evaluation restricts the manner in which treatment is assigned. For example, treatment may be randomly assigned. In an experimental context, the distinction between approaches to causal inference based on control and randomization dates back at least to Fisher (1935). The econometric approach to restricting the manner of treatment assignment is to impose an exclusion restriction:

Condition 2 *There is a random variable Z such that for all z*

$$E[Y_0|Z=z] = E[Y_0]$$

and

$$E[P|Z=z] \quad \textit{is a non trivial function of } z$$

The covariate Z affects the participation probability, but is not related to the expected response in the absence of treatment.

Exclusion restrictions are widely used in econometrics, usually in conjunction with other identifying restriction. One of the most influential approaches is that developed in a series

of papers by Heckman (1976, 1979). The following example is a simplification of the model used by Heckman (1979):

$$(3) \quad Y = P \cdot Y_1 + (1 - P) \cdot Y_0 = \mu + \alpha \cdot P + \varepsilon$$

$$(4) \quad P = I[\gamma \cdot Z + U \geq 0]$$

$$(5) \quad \begin{pmatrix} \varepsilon \\ U \end{pmatrix} \Big| Z \sim \mathcal{N}\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \Sigma \right)$$

The conditional expectation of Y_0 given $Z = z$ is

$$E[Y_0|Z = z] = \mu + E[\varepsilon|Z = z] = \mu$$

and since participation depends on both Z and U, it satisfies Condition 2. Notice that the treatment effect α in (3) is identical for every subject, so it is equal to the average and selected average treatment effects.

Another example is Angrist's (1991) nonlinear model with an omitted variable, U, that is correlated with P, but independent of an excluded instrument Z:

$$E[Y|P = p, U = u, Z = z] = F(p, u; \beta)$$

$$E[P|U = u, Z = z) = G(u, z; \gamma)$$

and Z and U independent. Angrist shows that the average treatment effect

$$\tilde{\alpha} = E[F(1, U; \beta) - F(0, U; \beta)]$$

is identified if and only if F or G is additively separable. In most of the econometric literature identification is based on distributional assumptions, functional form assumptions regarding either the conditional expectation of the response function and the probability of participation, or both. In our main result, we investigate when the exclusion restriction outlined in condition 2 is sufficient to identify the average treatment effect. Our approach is to invoke easily verifiable restrictions on the value of $h(z,u)$ and the distribution of Z.

Condition 3 *There is a set $\mathcal{Z}_0$ such that $1 > Pr(Z \in \mathcal{Z}_0) > 0$, $Pr(P = 1|Z = z) = 0$ for all $z \in \mathcal{Z}_0$.*

Theorem 1 *Conditions 2 and 3 are sufficient for identification of α with a random sample of (Y, Z, P).*

Proof: Let A be an indicator for the event $Z \notin \mathcal{Z}^0$. Then:

$$E[Y|A = 0] = E[Y_0]$$

$$E[Y|A = 1] = E[Y_0|A = 1] + Pr[P = 1|A = 1] \cdot E[Y_1 - Y_0|A = 1, P = 1]$$

$$= E[Y_0] + Pr[P = 1|A = 1] \cdot E[Y_1 - Y_0|P = 1]$$

Since we can consistently estimate $Pr[P = 1|A = 1]$, $E[Y|A = 0]$ and $E[Y|A = 1]$ we can identify $\alpha = E[Y_1 - Y_0|P = 1] = \{E[Y|A = 1] - E[Y|A = 0]\}/Pr(P = 1|A = 1)$.

$\mathcal{QED}$.

The theorem above shows that it is sufficient for identification to have a value, or set of values, Z_0, which is realized with non-zero probability and for which the probability of participation is zero. The question arises whether this is a necessary as well as a sufficient condition. A complete answer is difficult to give. But, a number of related results suggest that it is almost impossible to achieve identification otherwise. First we note that the key to identifying α is the identification of $E[Y_0]$:

Result 1 *α is identified if and only if $E[Y_0]$ is identified.*

Proof: By definition $\alpha = E[Y_1|P=1] - E[Y_0|P=1]$. Note that $E[Y_1|P=1]$ is identified because we observe Y_1 if $P=1$. Therefore identification of α is equivalent to identification of $E[Y_0|P=1]$. This is equal to $\{E[Y_0] - (1 - Pr(P=1)) \cdot E[Y_0|P=0]\}/Pr(P=1)$. Because $E[Y_0|P=0]$ and $Pr(P=1)$ are identified, identification of $E[Y_0|P=1]$ is equivalent to identification of $E[Y_0]$.

QED.

Second, we show that if Z is a discrete random variable, Condition 3 is indeed necessary for identification of $E[Y_0]$ and therefore for identification of α:

Result 2 *Suppose Z is a discrete random variable with K points of support. If $Pr(P = 1|Z = z_k) > 0$ for all k, then $E[Y_0]$ is not identified without additional restrictions.*

Proof: We can identify from the sampling design, for $k = 1, \ldots, K$,

$$E[Y|Z = z_k] = E[Y_0] + Pr(P = 1|Z = z_k) \cdot E[Y_1 - Y_0|Z = z_k]$$

There are K equations in $K+1$ unknowns. Therefore we cannot identify $E[Y_0]$ without some restriction on $E[Y_1 - Y_0|Z = z_k]$ if $Pr(P = 1|Z = z_k] > 0$ for all k. Note that one restriction such as equality of the conditional difference $E[Y_1 - Y_0|Z = z_k]$ for k_1 and k_0 is sufficient for identification of $E[Y_0]$.

QED.

The reason that Results 1 and 2 do not constitute a complete argument for sufficiency is that if Z is not discrete, it might be possible to identify $E[Y_0]$ in certain limiting cases, even when Condition 3 fails. In fact, this sort of "identification at infinity" is an underlying theme of a number of previous results on the identification of treatment effects.

2.2 COMPARISON WITH PREVIOUS IDENTIFICATION RESULTS. Conditions 2 and 3 and Theorem 1 are related to some recent results on semi-parametric identification. In latent index models like (3)–(5), if the disturbances are normal then there is clearly no set Z_0 such that the participation probability is zero for that set. This implies that we cannot estimate $E[Y_0] = E[Y|A = 0]$, the expected response for those who had zero probability of participating, so that identification cannot be based on Theorem 1. However, one might be able to estimate $E[Y_0]$ in the limit. One such approach is Condition B in Heckman's (1990) theorem on nonparametric identification of treatment effects in a latent-index sample selection model. Heckman requires the support of $\gamma \cdot Z$ in the latent index to be the real line. Therefore, there is a sequence of sets Z_n such that the probability of participating goes to zero in the limit. That is, there is a sequence of sets Z_n, such that for all sequences of real

numbers $\eta_n > 0$ and $\delta_n > 0$ converging to zero, $Pr(Z \in \mathcal{Z}_n) > \delta_n$, and $Pr(P = 1|Z = z) < \eta_n$ for all $z \in \mathcal{Z}_n$. If the limit $\lim_{n\to\infty} E[Y_0|P = 0, Z \in \mathcal{Z}_n] = E[Y_0]$ for all such sequences η_n and δ_n, then an estimate based on such a sequence can take the place of $E[Y_0|A = 0]$ in the proof of Theorem 1 and identification is still obtained. This is similar to an earlier result in Chamberlain (1986) regarding semi-parametric identification of censored regression models. But both Chamberlain (p. 205) and Heckman (p. 317) seem to feel that this sort of "identification at infinity" is not a very compelling foundation for inference. Chamberlain explores the possibility of imposing additional mild restrictions that would actually rule out this result.

Identification at infinity is unnatural in latent index models partly because many, if not most, regressors in economics have bounded support and are discrete. Most importantly, however, the latent index framework is usually motivated from a model of individual choice. Although the economic theory of discrete choice is well-developed and generally accepted, the details of empirical implementation are not. Identification at infinity requires not only covariates shifting choices but excluded from outcomes, but also a covariate-choice relationship that obeys additional restrictions without intrinsic behavioral or institutional content.

Both our Theorem 1 and previous results rely on exclusion restrictions and restrictions on the probability of participating for certain groups. Therefore, identification under Theorem 1 is similar to identification under the results of Chamberlain (1986) and Heckman (1990). One essential feature, however, distinguishes our approach from the traditional econometric

viewpoint: In Theorem 1, the main source of identifying information – the set of covariates for whom the probability of participation is zero – is obtained from the knowledge that the program was simply not offered to certain individuals or groups. A latent index framework in this case is unnnatural and unnecessary; with this sort of prior information there is no need to rely on limiting behavior.[6]

Secondly, we note that Manski (1990) presents similar results regarding identification of density functions in selection models without reference to a latent index framework. Manski's Corollary 2 (p. 30) shows that given certain level–set restrictions, nonparametric bounds on density functions coincide, and therefore the density function is identified, if and only if the probability of selection is one for some part of the population. Like Heckman and Chamberlain, however, Manski (p. 30) seems to feel that identifying with level–set restrictions is "rarely identifying in practice." Part of the reason for this is that while the results by Manski and Heckman give identification in principle, Chamberlain proves that the information bound can be zero for these models. Our approach requires that $Pr(Z \in \mathcal{Z}_0) > 0$, which implies that the treatment effect is estimable at rate $\sqrt{N}$.

Finally, Imbens and Angrist (1991) discuss identification of local average treatment effects. The local average treatment effect, α_{zw}, is the expected treatment effect for individuals who would change their participation status if their value of Z were changed from z to w. Identification of the local average treatment effect does not require the existence of a group

[6]The argument we make for identifying information from program eligibility rules is similar to that made informally in a recent paper by Moffit (1991).

with zero participation probability. However, the exclusion restriction must be strengthened and requires that both Y_0 and Y_1 are independent of Z. In addition the relation between the instrument and participation is restricted. The result in Theorem 1 shows that the existence of an ineligible group directly reduces the need for untestable conditions for identification of a meaningful average treatment effect.

Recent empirical examples of evaluations in this framework include the geographically randomized Educational Assistance Test Program (EATP) and Multiple Option Recruiting Experiment (MORE), in which different packages of veterans educational benefits were randomized over military recruitment stations (Fernandez [1982]). In the EATP and MORE, new benefit packages were not offered to a random subsample of stations. Examples of observational research where the source of identifying information is derived from institutions include Angrist's (1990) use of the draft lottery to estimate the labor market consequences of Vietnam–era military service, Angrist and Krueger's (1989) use of birthday–ordering to estimate the effects of World War II military service, and Angrist and Krueger's (1991) use of the interaction between compulsory school attendance laws and quarter of birth to estimate the effects of compulsory schooling on earnings.

As in most econometric applications, the examples listed above were implemented using statistical models with a constant treatment effect, so that the exclusion restrictions alone are sufficient for identification. But selected average treatment effects may also be identified in some of these cases. For example, in the compulsory schooling application, virtually all

students born in certain quarters were compelled to complete an additional year of schooling. Other students chose whether or not to continue in school; the treatment in this case is failure to complete an additional year of schooling. Likewise, in the Vietnam-era draft lottery, virtually all non-deferred men with low lottery numbers were drafted.[7]

2.3 Identification of the Average Treatment Effect.

If instead of the set $\mathcal{Z}_0$ we had a set $\mathcal{Z}_1$ such that $Pr(P = 1|Z = z) = 1$ for all $z \in \mathcal{Z}_1$, we would be able to identify the selected average non-treatment effect:

$$-\bar{\alpha} = E[Y_0 - Y_1|P = 0]$$

The selected average non-treatment effect measures how much non-participants gain (or lose) from not participating in the program. This identification result is obvious if we reverse what we call treatment and non-treatment. If there is both a set $\mathcal{Z}_0$ that satisfies Condition 3 and a set $\mathcal{Z}_1$ that satisfies the above condition we can identify the average treatment effect $\bar{\alpha}$. The three treatment effects are related by the following identity:

$$\bar{\alpha} = Pr[P = 1] \cdot \alpha + (1 - Pr[P = 1]) \cdot \bar{\alpha}$$

Intuition for why α is identified is apparent from the proof of Theorem 1: $E[Y_0]$ is identified in the sample where $Z \in \mathcal{Z}_0$, and $E[Y_1]$ is identified in the set where $Z \in \mathcal{Z}_1$. If the treatment effect is identical for everybody then $\alpha = \bar{\alpha} = \bar{\alpha}$. In general however, the treatment effects for participants and non-participants can be different, and in that case identification of the

[7] For the identification results of this paper to hold in the lottery example, deferment would have to be an ignorable covariate.

average treatment effect (ATE) requires stronger assumptions than does identification of the selected treatment effect (SATE).

3.1 ELIGIBILITY–RANDOMIZATION AS AN EXPERIMENTAL DESIGN. Social experiments can be based on the random assignment of eligibility allowing individuals to freely choose whether or not to participate. This is an alternative to experiments where randomization takes place at the later stage where individuals have already expressed a willingness to participate.[8] Let D be an indicator, equal to 1 if someone is willing to participate and zero otherwise. Suppose there is some characteristic, indicated by a binary variable, A, where only people with $A = 1$ are eligible for treatment. In addition, assume that the joint distribution of response Y_i and willingness to participate D does not depend on A. Formally, we can write

$$P = A \cdot D$$

and

$$(6) \quad f(Y_i, D|A = 1) = f(Y_i, D|A = 0) \quad \text{for } i = 0, 1$$

The condition that is required for identification of $E[Y_1 - Y_0|P = 1]$ using Theorem 1 is that $E[Y_0|A = 1] = E[Y_0|A = 0]$. Equation (6) is much stronger than this, but it makes the identification strategy and the difference between the two types of experiments transparent.

[8] An example of this type of design is the National Supported Work Demonstration, analyzed by Lalonde (1986).

A direct consequence of (6) is that $E[Y_0|A = a] = E[Y_0]$ and therefore A satisfies both conditions 2 and 3 and we can identify the selected average treatment effect

$$\alpha = E[Y_1 - Y_0|P = 1] = E[Y_1 - Y_0|D = 1] = \frac{E[Y|A = 1] - E[Y|A = 0]}{Pr(P = 1|A = 1)}.$$

The SATE is in this case the expected treatment effect for all participants if eligibility were to be extended to the entire population, i.e. if $A = 1$ for all individuals.[9] The combination of the ineligibles and eligible non-participants allows us to identify the distribution of Y_0 for those who are willing to participate:

$$f(Y_0|D = 1) = \frac{1}{Pr(D = 1)} f(Y_0|A = 0) - \frac{1 - Pr(D = 1)}{Pr(D = 1)} f(Y_0|A = 1, P = 0),$$

where $Pr(D = 1) = Pr(D = 1|A = 1)$ is identified from the proportion of participants among eligibles. One advantage of this type of experiment rather than the randomizing of applicants is that we also observe a number of individuals who do not wish to participate and therefore we can identify the selection effect β.

Harris (1985), Garfinkel, Manski and Michalopoulos (1991) and Moffit (1991) refer to experiments based on site randomization as macroexperiments, in contrast to microexperiments in which individuals within a site are randomly assigned to treatment and control groups. These authors stress that such macroexperiments can potentially identify macro treatment effects that result from interaction between individuals. An important difference

[9]Our result differs from that in Heckman (1990b, p. 27) because we compare eligibles and ineligibles whereas Heckman compares participants and ineligibles.

between our approach and previous discussion of macro experiments, however, is that we are not arguing for saturation of treatment within eligible sites.[10]

A further advantage of an experiment in which eligibility is randomly assigned is that there is no formal application process for subjects who will later be randomized out. The need to deny treatment appears to be major factor in the dissatisfaction of job training centers with randomized assigment (Manski and Garfinkel 1991). Morcoever, in medical research, eligibility randomization does not require that individual physicians deny a treatment they feel is beneficial (as ocurred in the controversial ECMO [extracorporeal membrane oxygenation] study of infant mortality; see Royall, [1991]). Instead of randomizing treatment within hospitals, randomly chosen hospitals could have been selected for study, with physicians freely choosing the most appropriate treatment within eligible sites, and data collected on outcomes at all sites. Another issue of interest in medical research is the question of non-compliance in conventional clinical trials. It is clear that as long as eligibility for treatment ("intention-to-treat" in biometric terminology) is randomized, the effect of a binary treatment on participants is identified using Theorem 1.

3.2 LINEAR INSTRUMENTAL VARIABLES ESTIMATION. In this section we show that if conditions 2 and 3 are satisfied we can estimate α in a straightforward manner. First we discuss the case where we observe Y, P and A, an indicator for the event $Z \notin \mathcal{Z}_0$. In this

[10]The Wisconsin Child Support Demonstration (Garfinkel 1983) follows this basic approach. Treatments in this study are randomized over Wisconsin counties, although the focus is on county-level outcomes and not the individual outcomes captured by SATE as defined here.

case we can estimate α by linear instrumental variables. Second, we analyze the case where we observe Y, P and Z. It turns out that this does not necessarily increase the efficiency of our estimate of the treatment effect. Finally we discuss estimation if we do not observe A itself, but a variable correlated with A.

The first estimator is a linear instrumental variables estimator. The variable A is an instrument for the endogenous regressor P because

$$(7) \quad E[Y|A] = E[Y_0] + E[P|A] \cdot E[Y_1 - Y_0|P = 1]$$

The sample analog of the solution for α is an estimate of $\text{Cov}(y, A)/\text{Cov}(P, A)$:

$$\hat{\alpha} = \frac{\bar{Y}_{A=1} - \bar{Y}_{A=0}}{\bar{P}_{A=1}}$$

where $\bar{Y}_{A=1}$ and $\bar{P}_{A=1}$ are sample averages conditional on $A = 1$.

The question naturally arises whether we can improve on this estimate of the selected average treatment effect if we observe Z as well as A. Using Chamberlain's (1990) approach to semi-parametric efficiency bounds one can show that this is only possible if Z affects the conditional variance of Y_0:

Theorem 2 *If the conditional variance of Y_0, $E[(Y_0 - E[Y_0])^2|Z = z]$ does not depend on z, then $\hat{\alpha}$ is, within the the class of regular estimators that only use the restrictions implied by Conditions 2 and 3, an efficient estimator for α. If the conditional variance does depend on z, one can obtain a more efficient estimator by replacing $\bar{Y}_{A=0}$ in the formula for $\hat{\alpha}$ by an efficient estimator for $E[Y_0]$ that adapts for the heteroscedasticity.*

Proof: see appendix.

Given A, there is only useful information in Z if Y_0 is conditionally heteroskedastic. However, it is unusual to have a case where one has a convincing argument that Z does not belong in the conditional mean function, but does belong in the conditional variance function. Therefore, in most cases the instrumental variables estimator based on A will be efficient.

Finally, note that if conditions 2 and 3 are satisfied but we do not observe A, we can still consistently estimate α if we observe a random variable X satisfying

Condition 4 $\quad E[Y|A,X] = E[Y|A] \qquad E[P|A,X] = E[P|A]$

and $E[P|X = x]$ *is a nontrivial function of* x.

Condition 4 implies that X affects mean outcomes and treatment probabilities only through its effect on eligibility A. In this case we can use X as an instrument instead of A. To see this, note that from (7);

$$E\Big\{E[Y|A]|X\Big\} = E[Y_0] + E[Y_1 - Y_0|P = 1] \cdot E\Big\{E[P|A]|X\Big\}$$

which simplifies to

$$E[Y|X] = E[Y_0] + E[Y_1 - Y_0|P = 1] \cdot E[P|X]$$

This implies that X is a valid instrument. It is clear that using both X and A as instruments is equivalent to using just A because X does not add any information once A is known. However, X may be useful if A is not observed. In the example of the draft lottery, one might envision knowing the week a person was born, but not the exact day. In that case the

week is the inaccurate instrument X while the actual day on which a person is born would be the accurate instrument A.

4. INFERENCE WHEN THE SELECTED AVERAGE TREATMENT EFFECT IS NOT IDENTIFIED. In this section we discuss what can be learned about treatment effects when Condition 2 is satisfied, but not Condition 3. There is no control group of ineligibles, and therefore an essential component of the instrumental variables approach discussed in the previous section is missing. We also assume that Z is discrete with points of support $z_1, z_2, \ldots, z_L$. Let $p_{z_k} = Pr(P = 1|Z = z_k)$, $\alpha_{z_k} = E[Y_1 - Y_0|P = 1, Z = z_k]$, $\pi_k = Pr(Z = z_k)$ and $Q = Pr(P = 1) = \sum \pi_k p_{z_k}$. In terms of these parameters, the selected average treatment effect is equal to

$$\alpha = \sum_k \frac{\pi_k \cdot p_{z_k}}{Q} \cdot \alpha_{z_k}$$

The probability limit of the instrumental variables estimator of the parameter α_λ in the equation

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

where $\varepsilon = Y_0 - E[Y_0]$ and the term in curly brackets is a compound error term. The instrument vector is $[1\ E[P|Z])$. This is the optimal instrument when the compound error term is conditionally (on Z) homoskedastic (Newey [1989]). Define

$$\lambda_k = \frac{p_{z_k}(p_{z_k} - Q)}{E[p_{z_k}(p_{z_k} - Q)]}$$

and

$$\lambda = \frac{E[p_{z_k}^2]}{E[p_{z_k}(p_{z_k} - Q)]} > 1$$

The λ_k are weights that have expectation equal to one, but they can be negative. Note that

$$E[Y|Z = z_k] = E[Y_0] + \alpha_{z_k} \cdot p_{z_k}$$

Then we have:

Result 3 *The instrumental variables estimator for α using Z as an instrument for P has probability limit*

$$\alpha_\lambda = E[\lambda_k \cdot \alpha_{z_k}] = \lambda \cdot \alpha_0 + (1 - \lambda) \cdot \alpha$$

where

$$\alpha_0 = E\frac{p_{z_k}^2}{E(p_{z_k}^2)}\alpha_{z_k}$$

Proof: The first part follows directly from the expression for $E[Y|Z = z_k]$ and the definitions for α_{z_k} and λ_k. To see the second part, write:

$$\frac{1}{1-\lambda} \cdot [\alpha_\lambda - \lambda\alpha_0]$$
$$= \frac{1}{1-\lambda}\left[\frac{E[p_{z_k}(p_{z_k} - Q)\alpha_{z_k}]}{E[p_{z_k}(p_{z_k} - Q)]} - \frac{E[p_{z_k}^2]}{E[p_{z_k}(p_{z_k} - Q)]} \cdot \frac{E[p_{z_k}^2]}{E[p_{z_k}^2]}\alpha_{z_k}\right]$$

which simplifies to $E[p_{z_k}\alpha_{z_k}/Q]$ which is equal to α. $\mathcal{QED}$.

α_0 is a weighted average treatment effect, with weights proportional to $p_{z_k}^2$. If the treatment effect is constant, then both α_0 and α_λ coincide with the selected average treatment

effect α. Therefore if we are prepared to bound the treatment effect heterogencity[11], we can calculate bounds for the selected average treatment effect as follows: Note that λ in the above result is estimable from the data.[12] Define $c = \alpha_0/\alpha$. In terms of c and λ the bias of the IV estimator is:

$$\frac{\alpha}{\alpha_\lambda} = \frac{1}{\lambda \cdot c + (1-\lambda)}$$

Given a choice of c, the bound is estimable.

An alternative to bounding the treatment effect heterogeneity is the approach in Imbens and Angrist (1991) where restrictions on the way in which the instrument Z affects participation is employed to identify a local average treatment effect.

5. CONCLUSION. The SATE measures the average difference between the outcomes of program participants and what participants' outcomes would have been had they not been treated. When some individuals or groups are ineligible to participate in a program, and eligibility does not affect outcomes for other reasons, the SATE is identified using a simple instrumental variables estimator. This estimator will usually be efficient – it makes full use of the identifying information provided by program eligibility rules.

The possibility of identification through eligibility rules is established using the same logic as recent arguments for identification based on the existence of a set of covariates for which the probability of treatment approaches zero in the limit. The source of identifying

[11]This is an alternative to bounds on the response variable itself, which is analyzed as in the context of selection models in Manski (1991b)

[12]In Angrist and Krueger (1990) λ is estimated to be about 2.5 for the relation between quarter of birth and high school graduation.

information is different, however, and likely to be more credible than identification through latent index models of individual behavior. Program rules are a matter of public record, and observed data can be used to verify enforcement of the rules. Identification through eligibility rules may also provide a good forecast of future program effects under the same rules. Another attractive feature of this approach is that no eligible participant need be denied treatment in experimental designs based on this principle.

Appendix

Proof of Theorem 2: The selected average treatment effect α is equal to $(E[Y|A = 1] - E[Y|A = 0])/(Pr(P = 1|A = 1)$. An efficient estimator can therefore be obtained by substituting efficient estimators for $E[Y|A = 1]$, $E[Y|A = 0]$ and $Pr(P = 1|A = 1)$ in this formula. We will show that

1) $\bar{Y}_{A=1}$ is an efficient estimator for $E[Y|A = 1]$,

2) $\bar{P}_{A=1}$ is an efficient estimator for $Pr(P = 1|A = 1)$,

3) $\bar{Y}_{A=0}$ is an efficient estimator for $E[Y|A = 0]$ if the conditional variance of Y_0 given Z does not depend on Z.

There are two steps. First we show that the model can be characterized by a finite number of conditional moment restrictions. Second we show that given those moment restrictions the three estimators are efficient.

The model implies the following conditional moment restrictions: If $A = 0$ then

$$E[Y - \theta|Z] = 0 \quad E[P|Z] = 0$$

If $A = 1$ then

$$E[P - h_1(Z)|Z] = 0 \quad E[Y - h_2(Z)|Z] = 0$$

The model does not imply any other restrictions. It is essential to show this before proving efficiency using the Chamberlain bounds The argument goes as follows. Suppose we have a datagenerating process for (Y, Z, P) with P binary, satisfying the moment conditions. Then we can always construct a model that satisfies (1) and (2) as follows:

$$E[Y_0|Z] = \theta$$

Choose any non–constant function for $E[Y_0|Z, P = 0]$, let

$$E[Y_0|Z, P = 1] = \{\theta - (1 - h_1(Z))E[Y_0|Z, P = 0]\}/h_1(Z)$$

and then we complete the model by choosing for all $Z \notin \mathcal{Z}_0$

$$E[Y_1|Z, P = 1] = \{h_2(Z) - \theta + h_1(Z) \cdot E[Y_0|Z, P = 1]\}/h_1(Z)$$

This constructed model satisfies $E[Y_0|Z] = 0$ for all Z. For this construction it is essential to have $h_1(Z) > 0$ which is true when $A = 1$ by definition.

Given that the model is fully characterized by the conditional moments, it is straightforward to derive the bounds for the three quantities of interest: $E[h_1(Z)]$, $E[h_2(Z)]$ and θ. The formulas in Chamberlain (1990, p 7) can be applied and simplified directly. First θ. Given a set of N_0 observations with $A = 0$, the bound on the variance of $\sqrt{N_0}(\hat{\theta} - \theta^*)$ is

$$\left\{E\left[\frac{1}{E[(Y_0 - \theta^*)^2|Z]}\right]\right\}^{-1}$$

This simplifies to $E[(Y_0-\theta^*)^2]$ if there is no heteroskedasticity. The variance of $\sqrt{N_0}(\bar{Y}_{A=0}-\theta^*)$ is $E[(Y_0-\theta^*)^2]$. Therefore $\bar{Y}_{A=0}$ is efficient if there is no heteroskedasticity. In exactly the same way we look at the variance bound for $E[h_1(Z)]$ and $E[h_2(Z)]$ given a set of N_1 observations with $A=1$. In both cases the variance is equal to the variance of the average. In other words, $\bar{Y}_{A=1}$ and $\bar{P}_{A=1}$ are efficient estimators.

$\mathcal{QED}$.

REFERENCES

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

ANGRIST, J.D., (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, (1991b), "Does compulsory school attendance affect schooling and earnings" Harvard Institute of Economic Research Discussion paper No. 1475, March.

BARNOW, B., CAIN, G., AND A. GOLDBERGER, (1981), "Selection on Observables." *Evaluation Studies Review Annals*, 5, 43–59, Bev Hills, Sage

CENTER FOR DRUG EVALUATION AND RESEARCH, (1988), "Guidline for the Format and Content of the Clinical and Statistical Sections of a New Drug Application", Food and Drug Administration, Department of Health and Human Services, Washington, D.C.: US-GPO.

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

—, (1990), "Efficiency Bounds for Semi-Parametric Regression", Harvard Institute of Economic Research Discussion Paper, No. 1494, June.

EFRON, B., AND D. FELDMAN, (1991), "Compliance as an Explanatory variable in Clinical Trials," Journal of the American Statistical Association 86 (March 1991), 9-17.

FERNANDEZ, R. L., (1982), "Enlistment Effects and Policy Implications of the Educational

Assistance Test Program," Santa Monica: Rand Corporation.

FISHER, R. A. (1935), "The Design of Experiments", Edinburgh: Oliver and Boyd.

GARFINKEL, I., (1983), Evaluation Design for Child Support Demonstration, University of Wisconsin Institute for Research on Poverty, Discussion Paper 714–82.

GARFINKEL, I., C. MANSKI AND C. MICHALOPOULOS, (1990), "Micro Experiments and Macro Effects", University of Wisconsin Insitute for Research on Poverty", Discussion Paper, 922-90.

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

HAM, J.C. AND R.J. LALONDE, (1990), "Using Social Experiments to Estimate the Effect of Training on Transition Rates," in Panel Data and Labor Market Studies, edited by J. Hartog, G. Ridder, and J. Theeuwes, Amsterdam: North- Holland.

HARRIS, J. E., (1985), "Macroexperiments versus Microexperiments for Health Policy", in: Social Experimentation, edited by J. A. Hausman and D. A. Wise, Chicago, University of Chicago Press.

HECKMAN, J. J., (1976), "The common structure of statistical models of truncation and censoring," *Annals of Economic and Social Measurment*, 5:475–492.

—, (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.

—, (1990b), "Randomization and Social Policy Evaluation", mimeo, Chicago University, Novem-

ber 1990.

—, AND BO HONORÉ, (1989), "The Empirical Content of the Roy Model," *Econometrica*, (September 1990).

—, 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.

HOTZ, J., (1991), "Recent Experience in Designing Evaluations of Social Programs: The Case of the National JPTA Study", in: Evaluating Welfare and Training Programs, edited by Manski and Garfinkel, Cambridge: Harvard University Press, forthcoming.

IMBENS, G., AND J. D. ANGRIST, (1991), "Identification and Estimation of Local Average Treatment Effects", Department of Economics, Harvard University.

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

MANSKI, C. F., (1990a), "Nonparametric Bounds on Treatment Effect", *American Economic Review*, 80(2), 319–323.

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

—, AND IRWIN GARFINKEL, (1991), Evaluating Welfare and Training Programs, edited by Manski and Garfinkel, Cambridge: Harvard university Press, forthcoming.

MOFFIT, R. (1991a), "Program Evaluation with Non–Experimental Data", forthcoming in *Evaluation Review*

—, (1991b), "Evaluation Methods for Program Entry Effects", in: Evaluating Welfare and Training Programs, edited by Manski and Garfinkel, Cambridge: Harvard University Press, forthcoming.

NEWEY, W., (1989), "Efficient Instrumental Variables Estimation of Nonlinear Models", Princeton University Econometric Research Memorandum No 341.

PERMUTT, THOMAS AND J. RICHARD HEBEL, (1989), "Simultaneous-Equation Estimation in a Clinical Trial of the Effect of Smoking on Birth Weight," *Biometrics*, 45 (June): 619-622.

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.

ROSENBAUM, P.R. AND D.B. RUBIN, (1983), "Assessing Sensitivity to an Unobserved Binary Covariate in an Observational Study with Binary Outcome," *Journal of the Royal Statistical Society Series B*, 45(2): 212-218.

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.

RUBIN, D., (1977), "Assigment of Treatment on the Basis of a Covariate," *Journal of Educational Statistics*, 2(10): 1-26