TECHNICAL WORKING PAPER SERIES

RANDOMIZATION AS AN INSTRUMENTAL VARIABLE

James J. Heckman

Technical Working Paper No. 184

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
September 1995

This research was supported by NSF-SES-93-34568 and a grant from the Russell Sage Foundation. I thank the editor, Robert Moffitt, for his comments, as well as an anonymous referee. 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 #184
September 1995

RANDOMIZATION AS AN INSTRUMENTAL VARIABLE

ABSTRACT

This paper discusses how randomized social experiments operate as an instrumental variable. For two types of randomization schemes, the fundamental experimental estimation equations are derived from the principle that experiments equate bias in control and experimental samples. Using conventional econometric representations, we derive the orthogonality conditions for the fundamental estimation equations. Randomization is a multiple instrumental variable in the sense that one randomization defines the parameter of interest expressed as a function of multiple endogenous variables in the conventional usage of that term. It orthogonalizes the treatment variable simultaneously with respect to the other regressors in the model and the disturbance term for the conditional population. However, conventional "structural" parameters are not in general identified by the two types of randomization schemes widely used in practice.

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

Randomized social experiments are now coming into use. Their limitations and benefits are now beginning to be understood. Papers by Burtless (1995), Heckman (1992), Heckman and Smith (1995a) and Moffitt (1992) among others clarify the behavioral and statistical assumptions underlying the experimental method.

This paper contributes to this literature and considers the social experiment as an instrumental variable. It develops the point that under the assumptions that justify its application, widely-used randomization schemes do not achieve their results by producing exogeneity of the treatment with respect to the population error term, as that term is ordinarily used in econometrics. Rather, these randomizations operate by balancing or equating the bias in the sample of persons randomized into a program with the bias in the sample of persons randomized out of the program. Randomization creates independence of the treatment effect with respect to other regressors and with respect to the error term in conditional populations. One randomization generates a multiple instrumental variable. Treatment effects as functions of an arbitrarily large number of endogenous variables can be identified from one randomization.

I develop these points for two distinct economic models: (a) a common effect model (treatment has the same effect on everyone with the same observed X characteristics) and (b) a variable effect model (treatment has different effects on everyone with the same observed X characteristics). The latter model is also known as a random effects model. The former is the one most widely used in applied work. Heckman and Robb (1985) and Heckman (1992) demonstrate the value in distinguishing between these two models in devising strategies for evaluating social programs.

I first consider randomization administered at the stage where persons apply to and are accepted into a social program and are then randomized out of the program. Randomization administered at that stage is widely used. Under the conditions specified in Heckman (1992) and Heckman and Smith (1995), this randomization identifies the mean gain to participating in the program for those who would usually participate in it. This mean gain is sometimes called the effect of treatment on the treated. I also consider samples produced by randomizing eligibility for the program.

I begin by briefly stating the evaluation problem.

1. The Evaluation Problem

The evaluation problem is a missing data problem. Persons may be in either one of two states but not both at the same time. The states are denoted "0" and "1" respectively. Outcomes are (Y_0, Y_1). Let $d = 1$ if a person is in state "1"; $d = 0$ otherwise. The outcome observed for an individual, Y, is

$$Y = dY_1 + (1 - d) Y_0.$$

This is an instance of the Roy model (1951) or a switching model (see Goldfeld and Quandt, 1972). Statisticians call this the "Rubin model" after one clear exposition of Fisher's model of experiments set forth by Rubin (1978). If "0" is the no program state, and "1" is the program state, the gain to participating in the program is

$$\Delta = Y_1 - Y_0.$$

If, contrary to hypothesis, we could simultaneously observe Y_1 and Y_0 for the same person, there would be no evaluation problem. One could construct Δ for everyone.

To cast the model into familiar econometric notation, write the two population model of Fisher (1951), Cox (1958), Roy (1951) or Rubin (1978) as a function of observables (X) and unobservables (U_1, U_0):

1(a) $$Y_1 = g_1(X) + U_1$$

1(b) $$Y_0 = g_0(X) + U_0$$

where

$$E(U_1) = 0 = E(U_0).$$

It is assumed that g_1 and g_0 are nonstochastic functions. For the familiar case of linear regression,

$$g_1(X) = X\beta_1$$

$$g_0(X) = X\beta_0.$$

There are many forms of the evaluation problem depending on what feature of the missing data one seeks to construct. The most common form of the problem is cast in terms of means. One mean receives the most attention:

The Mean Effect of Treatment on The Treated

(2) $$E(Y_1 - Y_0 \mid X, d = 1) = E(\Delta \mid X, d = 1) = g_1(X) - g_0(X) + E(U_1 - U_0 \mid X, d = 1).$$

This mean answers the question "how much did persons participating in the program benefit compared to what they would have experienced without participating in the program?". It is a non-standard parameter from the vantage point of conventional econometrics because it combines "structure" (the g_0 and g_1 functions) with the means of error terms (U_0 and U_1).[1]

A second mean also receives some attention in the literature - the effect of randomly selecting persons from the general population into the program:

Mean Effect of Treatment Randomly Applied To The Population

(3) $$E(Y_1 - Y_0 \mid X) = E(\Delta \mid X) = g_1(X) - g_0(X) + E(U_1 - U_0 \mid X).$$

This mean answers the question of how much the average outcome would be affected if participation in a program were universal, assuming that there are no general equilibrium effects. Alternatively, this parameter is the effect of taking a person from the general population at random and moving him or her from "0" to "1". Further discussion of these parameters and their relationship to the traditional parameters of cost-benefit analysis is presented in Heckman and Smith (1995b).

Although the assumption of separability between X and U is conventional in econometrics, it is not required to define $E(Y_1 - Y_0 \mid d = 1, X)$ or $E(Y_1 - Y_0 \mid X)$ nor is it necessary to assume such separability in deriving estimates from experiments.

For certain purposes, it is also of interest to inquire about distributions of gains:

4(a) $$F(\Delta \mid d = 1, X)$$

or

4(b) $$F(\Delta \mid X)$$

but it has been shown that social experiments, unaccompanied by further assumptions, cannot recover these distributions (see Heckman, 1992, or Clements, Heckman and Smith, 1993). Under ideal conditions, social experiments recover

$$F(Y_0 \mid d = 1, X) \text{ and } F(Y_1 \mid d = 1, X)$$

[1]Heckman and Robb (1985, 1986) present conditions for identifying this parameter using instrumental variables in nonexperimental settings. Their conditions apply to the general "variable treatment effect case" of equations 1(a) and 1(b).

if randomization is administered at a stage of the application and acceptance decision at which persons would ordinarily be accepted into programs, and there is no attrition from the program.

The evaluation problem arises from the fact that ordinary observational data do not provide sample counterparts for the missing counterfactuals. For means, experiments supplement observational data by providing the information needed to form the sample counterpart of

$$E(Y_0 \mid d = 1, X).$$

More generally, social experiments supplement observational data and produce the information needed to form the empirical distribution counterpart of

$$F(Y_0 \mid d = 1, X).$$

Randomization provides the sample counterparts to these population objects if randomization bias induced by the process of experimentation is assumed to be unimportant. (Heckman, 1992).

2. *Randomization Balances Bias*

First consider randomization at the stage where persons apply to and are accepted into a social program. Nowhere is it assumed that $E(U_1 \mid X) = 0$ or $E(U_0 \mid X) = 0$. Thus X can fail to be exogenous in the conventional sense of that term. Yet randomized trials that do not disrupt the program, and are not subject to attrition or non-compliance, produce the data that can be used to consistently estimate parameter (2). This highlights both the unusual nature of that parameter and the benefits of randomization.

To establish how randomization identifies (2), it is instructive to introduce new variables denoted by *. $d^* = 1$ denotes the event: "in the presence of randomization a person would have participated in the program except possibly for being randomized out." We may also define Y_1^* and Y_0^* to be the outcomes observed under a regime of randomization. Absence of randomization bias for mean gain is defined as

(A-1) $$E(Y_1 \mid d = 1, X) = E(Y_1^* \mid d^* = 1, X)$$

and $$E(Y_0 \mid d = 1, X) = E(Y_0^* \mid d^* = 1, X)$$

i.e. randomization does not alter the outcome mean gain for the program being evaluated for all values of X.

Randomization operates conditionally on $d^* = 1$. This is appropriate because parameter (2) is defined conditionally. In the subpopulation for which $d^* = 1$, randomization operates by selecting persons into the program by a random device. $R = 1$ if a person is randomized into the program; $R = 0$ if a person is randomized out of the program. I assume that if $R = 1$, persons accept admission into the program and if $R = 0$, they do not obtain program services.

As a consequence of assumption (A-1), and the additional assumption that equates $R = 1$ or $R = 0$ with receipt of program services, using the definition $Y = Y_1 d + Y_0(1 - d)$,

5(a) $E(Y \mid d^* = 1, R = 1, X) = E(Y_1 \mid d = 1, X) = g_1(X) + E(U_1 \mid d = 1, X)$

5(b) $E(Y \mid d^* = 1, R = 0, X) = E(Y_0 \mid d = 1, X) = g_0(X) + E(U_0 \mid d = 1, X)$.

Randomization creates data that can be used to estimate counterfactual 5(b). Conditional mean 5(a) can be consistently estimated using ordinary observational data. If there is no randomization bias, and R is synonymous with receipt of services, both experiments and observational data are equally informative about 5(a). I assume no randomization bias and for notational simplicity henceforth equate d and d^*, Y_1 and Y_1^*, and Y_0 and Y_0^*.

Subtract 5(b) from 5(a) to obtain

$$E(Y \mid d = 1, R = 1, X) - E(Y \mid d = 1, R = 0, X) = g_1(X) - g_0(X) + E(U_1 - U_0 \mid d = 1, X) = E(\Delta \mid d = 1, X). \tag{6}$$

This can be consistently estimated using sample counterparts to population means. A nonparametric kernel estimator can be constructed using conventional methods. (See, e.g. Härdle, 1990). Nowhere is it necessary to assume that

$$E(U_1 \mid X) = 0 \quad \text{or} \quad E(U_0 \mid X) = 0.$$

In fact, it is clear from the definition of Δ that in general

$$E(U_1 - U_0 \mid d = 1, X) \neq 0.$$

To place randomization into a more familiar-looking instrumental variables framework, define $\tilde{U}_1$ as U_1 conditional on $d = 1$ and X and define $\tilde{U}_0$ as U_0 conditional on $d = 1$ and X. Then Y conditional on $d = 1$ and X may be written as $\tilde{Y}$, and

$$\tilde{Y} = [g_1(X) - g_0(X)]R + \{\tilde{U}_1 - \tilde{U}_0\}R + \tilde{U}_0. \tag{7}$$

In this notation, it is not assumed that $E(\tilde{U}_0 \mid d = 1, X) = 0$

nor is it assumed that $E(\tilde{U}_1 \mid d = 1, X) = 0$.

From definition (2),

$$\tilde{Y} = g_0(X) + E(\Delta \mid X, d = 1)R$$
$$+ \{\tilde{U}_0 + R(\tilde{U}_1 - \tilde{U}_0 - E(\tilde{U}_1 - \tilde{U}_0 \mid d = 1, X))\}.$$

Notice that the second component in braces has zero conditional mean by construction (the conditional mean of $\tilde{U}_1 - \tilde{U}_0$ is subtracted from $\tilde{U}_1 - \tilde{U}_0$).

Observe that as a consequence of randomization, R is independent of $g_0(X)$, $\tilde{U}_0$, and $\tilde{U}_0 - \tilde{U}_1 - E(\tilde{U}_1 - \tilde{U}_1 \mid d = 1, X)$ i.e.

$$R \perp\!\!\!\perp (g_0(X), \tilde{U}_0, \tilde{U}_1 - \tilde{U}_0 - E(\tilde{U}_1 - \tilde{U}_0 \mid d = 1, X)).$$

This defines the crucial set of orthogonality (really independence) conditions produced by an experiment. Since R is orthogonal with respect to $g_0(X)$, the non-orthogonality of $g_0(X)$ with respect to $\tilde{U}_0$ is irrelevant for identification of $E(\Delta \mid X, d = 1)$.

Formally, we may write out the orthogonality condition as

(8) $$E([\tilde{U}_0 + R(\tilde{U}_1 - \tilde{U}_0 - E(\tilde{U}_1 - \tilde{U}_0 \mid d = 1, X))] \mid R) = 0.$$

There is no assurance, however, that a similar condition holds for $g_0(X)$. In general

(9) $$E(g_0(X)[\tilde{U}_0 + R(\tilde{U}_1 - \tilde{U}_0 - E(\tilde{U}_1 - \tilde{U}_0 \mid X, d = 1))] \neq 0.$$

Because of the independence between R and $g_0(X)$ created by the experiment, (8) is sufficient to identify $E(\Delta \mid X, d = 1)$. Experiments do not in general identify $g_0(X)$, since in general $E(\tilde{U}_0 \mid X, d = 1) \neq 0$. Thus experiments of the type discussed in this section do not in general identify the structural parameters of the original equation but they identify parameter (2) provided that (A-1) is valid and persons assigned to treatment receive it and persons denied treatment do not. This point is obvious once it is recognized that the randomization considered in this section is conditional on a variable that is, in general, endogenous.

These expressions simplify when there is a common effect model. In that case $U_1 = U_0$, and $\tilde{U}_0 = \tilde{U}_1 \overset{def}{=} \tilde{U}$.

$$E(\Delta \mid X, d = 1) = g_1(X) - g_0(X)$$

and equation (7) may be written as

$$\tilde{Y} = g_0(X) + E(\Delta \mid X, d = 1)R + \tilde{U}.$$

Orthogonality condition (8) becomes

$$E(\tilde{U}_0 \mid R) = 0.$$

Again notice that in general

$$E(\tilde{U} g_0(X) \mid R) \neq 0.$$

A familiar form of the common effect model writes

$$g_0(X) = X\beta_0$$
$$g_1(X) = X\beta_1$$

so

$$E(\Delta \mid X, d = 1) = X(\beta_1 - \beta_0)$$

where the conditioning on d = 1 is often left implicit. An even more familiar form of the model writes

$$g_0(X) = X\beta_0$$
$$g_1(X) = X\beta_0 + \alpha.$$

This is the dummy endogenous variable model (Heckman, 1978). In this case

(10) $$\tilde{Y} = X\beta_0 + \alpha R + \tilde{U}.$$

Randomization ensures that R is independent of both $\tilde{U}$ and X. It does not ensure that X is independent of $\tilde{U}$. The orthogonality between R and X induced by an experiment implies that any dependence between X and $\tilde{U}$ does not affect the identifiability of α. Randomization creates an orthogonal regressor model for the subpopulation defined conditional on d = 1. Since R is independent of X and $\tilde{U}$, α is identified even if X is not orthogonal to $\tilde{U}$.

3. Discussion

Observe that one randomization identifies an entire function $E(\Delta \mid X, d = 1)$ over the support of X. In principle, $E(\Delta \mid X, d = 1)$ can be an infinite-dimensional function of X. Hence randomization is a multiple instrumental variable.

Observe further that randomization enriches the support of X in the following way. Suppose in the population that

$$\text{Support } (X \mid d = 1) \neq \text{Support } (X \mid d = 0).$$

Then in the subset of X values for which there is no overlap, observational methods cannot obtain comparisons for all X values and $E(\Delta \mid X, d = 1)$ cannot be identified for all X. Randomization creates a balanced support set

$$\text{Support}\ (X \mid d = 1, R = 1) = \text{Support}\ (X \mid d = 1, R = 0).$$

Unless $\text{Support}\ (X \mid d = 1) \subset \text{Support}\ (X \mid d = 0)$,

randomization enlarges the support set over which $E(\Delta \mid X, d = 1)$ can be defined and estimated. However, unless $0 < \Pr(d = 1 \mid X) < 1$, randomization does not identify $E(\Delta \mid X, d = 1)$ for all possible values of X. (See, e.g., Rosenbaum and Rubin, 1983). An extreme example of the benefit of randomization in enlarging the support set occurs when for certain values of X, the event $d = 0$ does not occur i.e.

$$\Pr(d = 0 \mid X) = 0$$

but $d = 1$ occurs with positive probability for all values of X:

$$0 < \Pr(d = 1 \mid X) < 1.$$

In this case, randomization expands the support of X given $d = 1$ to the entire support of X. It permits identification of $E(\Delta \mid X, d = 1)$ for all possible values of X.

Observe that in general, experiments defined conditional on $d = 1$, do not identify $E(\Delta \mid X)$, the effect of selecting a person at random from "0" and moving the person to "1". However, if the common effect model is assumed ($U_0 = U_1$), experiments conducted on populations defined conditional on $d = 1$ recover

$$E(\Delta \mid X),$$

the effect of selecting a person from the general population and placing them in the program. For in that case,

$$E(\Delta \mid X) = E(\Delta \mid d = 1, X).$$

(See Heckman, 1992 or Heckman and Smith (1995)).

One case where responses to treatment are heterogeneous, randomization is administered to those for whom d would have been "1" in the absence of randomization and the parameter (3) is nonetheless identified is as follows. Suppose at the time agents enroll in the program they forecast their gain to be the total population mean gain. Then clearly (2) equals (3) and the experiment identifies both parameters. (Heckman and Robb, 1985).

However, in general, if $U_0 \neq U_1$, $E(\Delta \mid X) \neq E(\Delta \mid d = 1, X)$.

4. Randomization of Eligibility

Randomization of eligibility for a program is sometimes proposed as a less disruptive alternative to randomization of admission among accepted applicants. (Heckman, 1992, Heckman and Smith (1993), Angrist and Imbens, 1991). In this section, I show that this type of randomization can be placed in an instrumental variable framework. Consider a population of persons ordinarily eligible for a program. For simplicity, this conditioning is kept implicit. Let e = 1 if a person is kept eligible after randomization; e = 0 if the person loses eligibility. Assume that assignment to eligibility does not disturb the underlying stochastic structure of (Y_0, Y_1, d, X) and that it is independent with respect to the outcome measures:

(A-2) $$(Y_1, Y_0, d, X) \perp\!\!\!\perp e,$$

and assuming $P(d = 1 \mid X) \neq 0$,

(11) $$\frac{E(Y|e=1,X) - E(Y|e=0,X)}{P(d = 1|X)} = E(\Delta \,|d=1,X) .$$

To prove this use the law of iterated expectations to obtain

12(a) $$\begin{aligned} E(Y \mid e = 1, X) = {} & E(Y_1 \mid d = 1, e = 1, X)\, P(d = 1 \mid e = 1, X) \\ & + E(Y_0 \mid d = 0, e = 1, X)\, P(d = 0 \mid e = 1, X) \end{aligned}$$

and

12(b) $$\begin{aligned} E(Y \mid e = 0, X) = {} & E(Y_0 \mid d = 1, e = 0, X) P(d = 1 \mid e = 0, X) \\ & + E(Y_0 \mid d = 0, e = 0, X) P(d = 0 \mid e = 0, X). \end{aligned}$$

From (A-2)

$$P(d = 1 \mid e = 1, X) = P(d = 1 \mid e = 0, X) = P(d = 1 \mid X)$$

so that the result follows by subtracting 12(b) from 12(a) and dividing by $P(d = 1 \mid X)$ provided $P(d = 1 \mid X) \neq 0$. Replacing population moments by sample moments, (11) is a version of Bloom's estimator for attrition from a program. (See Angrist and Imbens, 1991 and Heckman, Smith and Taber, 1995 for a discussion of Bloom's estimator).

I now present an instrumental variables interpretation of this estimator. Using equations 1(a) and 1(b), the law of iterated expectations and (A-2), one obtains

$$\begin{aligned} Y = {} & g_0(X) + e[g_1(X) - g_0(X) + E(U_1 \mid d = 1, X) - E(U_0 \mid d = 1, X)]P(d = 1 \mid X) \\ & + U_0 + e[U_1 - U_0 - [E(U_1 \mid d = 1, X) - E(U_0 \mid d = 1,X)]]P(d = 1 \mid X)] \end{aligned}$$

or

(13) $Y = g_0(X) + e[E(\Delta \mid d = 1, X)]P(d = 1 \mid X)$

$$+ \{U_0 + e[U_1 - U_0 - [E(U_1 \mid d = 1, X) - E(U_0 \mid d = 1, X)]]P(d = 1 \mid X)]\}.$$

Observe that by random assignment of e,

$$e \perp\!\!\!\perp \{U_0 + e[(U_1 - U_0 - [E(U_1 \mid d = 1, X) - E(U_0 \mid d = 0, X)]]P(d = 1 \mid X)\}$$

so that orthogonality (really independence) is an immediate consequence of the randomization although $g_0(X)$ need not be independent or orthogonal to the composite disturbance inside the braces of equation (13). Again, there is no requirement that X be independent or orthogonal with respect to U_1 or U_0. Hence, under standard rank conditions one can consistently estimate $E(\Delta \mid d = 1, X)P(d = 1 \mid X)$. Assuming that one can consistently estimate $P(d = 1 \mid X)$, one can estimate $E(\Delta \mid d = 1, X)$, the effect of treatment on the treated, by dividing the IV estimator of the product of the two terms by $P(d = 1 \mid X)$. Note that this randomization identifies $E(\Delta \mid d = 1,X)$ but not $E(\Delta \mid X)$.

5. Concluding Remarks

This paper considers randomization as an instrumental variable. Two types of randomizations are considered: (a) randomization of eligibility for a program and (b) randomization of admission into the program among eligible persons who would ordinarily be admitted into the program. The second type of randomization is widely used in conducting social experiments. Using a conventional separable-in-the-errors representation of equations, I have shown the orthogonality conditions that are produced by these two types of randomization schemes and how they identify a central parameter in program evaluation studies -the effect of treatment on the treated - parameter (2). One randomization serves to identify this parameter as a function of multiple endogenous variables as conventionally defined in econometrics.

Balancing the bias in experimental and control samples is the fundamental source of identification from experiments. Such balancing in no way depends on separability of errors from equations as conventionally assumed in econometrics (as in equations 1(a) and 1(b)) nor does it require that the X be either independent or orthogonal with respect to the U. The method of moments analogs to (6) or (10) can be implemented nonparametrically. These are the basic estimating equations for experiments from which the orthogonality conditions of this paper have been derived.

The fact that parameter (2) is not conventional has been the source of some confusion. It combines both structural portions (the g_0 and g_1 of equation 1(a) and 1(b) respectively) with conditional means of the errors (the U_1 and U_0). Experiments conducted at a stage where persons would ordinarily enter a program are not designed to consistently estimate the g_0 and g_1 functions and in general they do not. Experiments make the treatment variable orthogonal to the error and the other regressors thus separating the estimation of treatment effects from the estimation of the other parameters of the model.

Only under special conditions does either type of randomization discussed in this paper identify parameter (3) - the effect of moving a randomly selected person in the general population from state "0" to state "1". This is an intrinsically more difficult parameter to estimate because in most societies people cannot be forced to participate in programs against their will. It is more difficult to estimate $E(Y_1 \mid d = 0, X)$ than $E(Y_0 \mid d = 1, X)$ if responses to treatments are heterogeneous, and are partly anticipated at the time decisions to enroll in the program are made. By the same token, if there is attrition from the program, it may also be difficult to estimate (2) except under special conditions discussed in Bloom (1984), Heckman, Taber and Smith (1995), or Hotz and Sanders (1994).

Other types of randomization might be used. For example, if interest centers on estimating

$$Y = g(X) + U$$

and X is not independent of U, X might be experimentally varied as in the negative income tax experiments or in the electricity experiments. Then experimental variation in X can be used in the conventional way to produce an instrument for an endogenous variable. See the discussion in Heckman (1992).

References

Angrist, Joshua and Guido Imbens, "Sources of Identifying Information in Selection Models," NBER Technical Working Paper 117, 1991.

Bloom, Howard, "Accounting For No-Shows in Experimental Evaluation Designs", *Evaluation Review*, Vol. 82, #2, 1984, 225-246.

Burtless, Gary, "The Case For Randomized Field Trials in Economic and Policy Research", *Journal of Economic Perspectives*, Vol 9, 63-84, June 1995,

Clements, Nancy, James Heckman and Jeffrey Smith, "Making The Most Out of Social Experiments". First draft under revision, *Review of Economic Studies*, 1993.

Cox, David, The Planning of Experiments, New York, Wiley, 1958.

Fisher, Ronald, The Design of Experiments, 6th edition, Oliver and Boyd: London, 1951.

Goldfeld, Steven and Richard Quandt, Nonlinear Methods In Econometrics, North Holland: Amsterdam, 1972.

Härdle, Wolfgang, Applied Nonparametric Regression, Cambridge, England, Cambridge University Press, 1990.

Heckman, James, "Randomization and Social Program," in C. Manski and I. Garfinkle, eds., Evaluating Welfare and Training Programs, Cambridge, MA. Harvard University Press, 1992.

Heckman, James and Richard Robb, "Alternative Methods For Evaluating The Impact of Interventions," in Longitudinal Analysis of Labor Market Data, New York, Wiley, 1985.

_____, "Alternative Methods For Solving The Problem of Selection Bias in Evaluating The Impact of Treatments on Outcomes", in Howard Wainer, editor, Drawing Inferences from Self-Selected Samples, Berlin, Springer Verlag, 1986.

Heckman, James, Jeffrey Smith and Christoper, Taber, "Accounting for Dropouts in Evaluations of Social Experiments", under revision, *Review of Economics and Statistics*, 1994.

Heckman, James and Jeffrey Smith, "Assessing The Case For Randomized Evaluations of Social Programs", in K. Jensen and P.K. Madsen, eds., Measuring Labour Market Outcomes, Ministery of Labor, Copenhagen, Denmark, 1993.

_____, "Assessing The Case For Social Experiments", *Journal of Economic Perspectives*, Vol. 9, 84-110, June, 1995.

_____" Evaluating The Welfare State", Frisch Symposium, Oslo, Norway, March, 1995b.

Hotz, V. Joseph and Seth Sanders, "Bounding Treatment Effects in Controlled and Natural Experiments Subject to Post-Randomization Treatment Choice", Population Research Center, University of Chicago, 1994.

Moffitt, Robert "Evaluation of Program Entry Effects," in C. Manski and I. Garfinkle, eds., Evaluating Welfare and Training Programs, Cambridge, MA. Harvard University Press, 1992.

Rosenbaum, Paul and Donald Rubin, "The Central Role of the Propensity Score in Observational Studies For Causal Effects", *Biometrika*, Vol. 70, 1984, 41-55.

Rubin, Donald, "Bayesian Inference For Causal Effects: The Role of Randomization", *Annals of Statistics*, Vol. 6, #1, 1978, 34-58.

Roy, Andrew D., "Some Thoughts on The Distribution of Earnings", *Oxford Economics Papers*, Vol. 3, 1951, 135-146.