NBER WORKING PAPER SERIES

CHOOSING AMONG ALTERNATIVE NONEXPERIMENTAL METHODS FOR ESTIMATING THE IMPACT OF SOCIAL PROGRAMS: THE CASE OF MANPOWER TRAINING

James J. Heckman

Working Paper No. 2861

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
February 1989

James J. Heckman is the A. Whitney Griswold Professor of Economics, Department of Economics, Yale University, New Haven, CT 06520; and V. Joseph Hotz is Associate Professor, The Graduate School of Public Policy Studies, University of Chicago, Chicago, IL 60637. This research was supported by NSF grant number SES-87-39151 and a contract from the U.S. Department of Labor. The authors wish to thank Burt Barnow, Ricardo Barros, Kyungsoo Choi, Robert LaLonde, Ernst Stromsdorfer, and the referees for helpful comments. We thank Steve Cameron for able research assistance. We also thank Rebecca Maynard of MPR, Inc. for giving us her data. This paper is part of NBER's research program in Labor Studies. Any opinions expressed are those of the authors not those of the U.S. Department of Labor, the National Science Foundation, or the National Bureau of Economic Research.

NBER Working Paper #2861
February 1989

CHOOSING AMONG ALTERNATIVE NONEXPERIMENTAL METHODS FOR ESTIMATING THE IMPACT OF SOCIAL PROGRAMS: THE CASE OF MANPOWER TRAINING

ABSTRACT

The recent literature on evaluating manpower training programs demonstrates that alternative nonexperimental estimators of the same program produce a array of estimates of program impact. These findings have led to the call for experiments to be used to perform credible program evaluations. Missing in all of the recent pessimistic analyses of nonexperimental methods is any systematic discussion of how to choose among competing estimators.

This paper explores the value of simple specification tests in selecting an appropriate nonexperimental estimator. A reanalysis of the National Supported Work Demonstration Data previously analyzed by proponents of social experiments reveals that a simple testing procedure eliminates the range of nonexperimental estimators that are at variance with the experimental estimates of program impact.

James J. Heckman
Department of Economics
Yale University
P.O. Box 1972, Yale Station
New Haven, CT 06520

V. Joseph Hotz
The Graduate School of Public Policy Studies
University of Chicago
1155 East Sixtieth Street
Chicago, IL 60637

1. INTRODUCTION

Evaluation is now an accepted component of most government programs. Federal government support for evaluating social programs has ranged between 500 million and a billion dollars per year for the past decade. Rossi and Freeman (1985) describe a sampling of such studies. Manpower training programs, designed to upgrade the employment and earnings of the poor, have been a frequent subject of evaluation.

Virtually all of these manpower training evaluations are based on the following principle. Earnings and other outcome measures of trainees are compared with earnings and other outcome measures of nontrainees. In an experimental evaluation, the comparison group consists of individuals who applied and were accepted into the program but were randomized out before the program began. A comparison group selected in this fashion is sometimes called a control group. In a nonexperimental evaluation, the comparison group consists of individuals judged to be "comparable" to trainees except for not having received training. A variety of matching and statistical adjustment procedures have been proposed to account for discrepancies in observed and unobserved characteristics between trainees and candidate comparison group members that might distort earnings comparisons. Failure to control for such characteristics in comparing outcomes of trainees with comparison group members may lead to substantial bias in the estimate of program impacts. Such bias is called selection bias.

Nonexperimental estimators differ in the assumptions they invoke to justify various statistical adjustments. Precisely because data from a properly conducted experiment do not require such adjustments, experimental estimates of program impact are inherently less controversial.

In recent influential papers, several authors have argued that alternative nonexperimental estimators of manpower training program impacts produce a wide range of estimates for the same program. LaLonde (1986) and

Fraker and Maynard (1984, 1987) use experimental data from the National Supported Work Demonstration (NSW) project combined with nonexperimental data to compare experimental and nonexperimental estimates of the program. These authors find that nonexperimental estimates vary widely and differ greatly from the experimental estimates. Their findings have led respected scholars to conclude that

> "...estimates of program effects that are based on nonexperimental comparisons can be subject to substantial misspecification uncertainty" (Burtless and Orr, 1986, p. 613)

and that

> "...randomized clinical trials are necessary to determine program effects" (Ashenfelter and Card, 1985, p. 648).

Barnow (1987) argues that

> "...experiments appear to be the only method available at this time to overcome the limitations of nonexperimental evaluations" (p. 190).

The LaLonde and Fraker and Maynard studies stimulated the Department of Labor to fund a 20 million dollar evaluation of the Job Training Partnership Act using an experimental approach (see the recommendations of the Job Training Longitudinal Survey Research Advisory Panel in Stromsdorfer, et al. (1985).)

As noted by Rivlin (1971), randomized assignment of applicants is politically unpopular. There is considerable difficulty in gaining acceptance for such randomization by training program managers and local political officials. Moreover, as noted by Burtless and Orr (1986), there are practical difficulties in conducting experiments in social contexts. Individuals assigned to treatments often do not show up. Individuals randomized out of a program at one trial may subsequently cross over and become program participants. Experimental evaluation of multistage programs requires costly multistage randomization (Heckman, Hotz, and Dabos, 1987, p. 423). These difficulties compromise the ability of experiments to provide unbiased estimates of the impact of programs without resorting to the nonexperimental statistical adjustment methods that experiments were designed to avoid. Such costs make it likely that most evaluations will continue to be conducted on nonexperimental data. In light of the conclusions of the recent literature, however, it would seem that such evaluations are unlikely to be credible.

Two unstated premises underlie the recent negative assessments of nonexperimental evaluation procedures. The first premise is that alternative nonexperimental estimation procedures should produce approximately the same program estimate. The second premise is that there is no objective way to choose among alternative nonexperimental estimators.

The first premise is invalid if there are systematic differences between trainee and comparison group members in observed and unobserved characteristics affecting outcome measures. Different nonexperimental estimators make different assumptions about the distribution of these differences. Only in the absence of systematic differences in characteristics between trainee and comparison group members would alternative nonexperimental estimators produce the same estimate of program impact up to sampling variation. Evidence of striking differences in estimates produced from alternative nonexperimental estimators merely confirms the existence of systematic differences between trainees and comparison group members in characteristics affecting outcome measures. It sheds no light on the credibility of any particular estimator or class of estimators.

The truth of the second premise hinges on the content of the available nonexperimental data and on the assumptions invoked to justify various nonexperimental estimators. Below we present a test for model specification that can be applied to any nonexperimental model provided that there is access to pre-program outcome measures and regressor variables for trainees and comparison group members. Such data are widely available. We also present tests based on "overidentifying" aspects of certain evaluation models which impose restrictions on the data beyond those required to estimate program impacts. With access only to nonexperimental data of this type, one cannot test just-identifying assumptions associated with alternative nonexperimental models. We go on to present specification tests that exploit less commonly available experimental data.

This paper demonstrates the value of such tests for choosing among alternative nonexperimental estimators. Previous empirical studies documenting the sensitivity of estimates of program impacts to alternative nonexperimental estimation procedures do not test the fitted models against the available data or else disregard the inference from such tests.

We reanalyze the NSW data used by LaLonde (1984) and Fraker and Maynard (1984, 1987) in their influential studies. We demonstrate the value of our proposed tests in rejecting nonexperimental estimators that produce estimates in discord with the experimental evidence and in not rejecting estimators that produce estimates that accord with the experimental evidence. Our evidence tempers the recent pessimism about nonexperimental evaluation procedures that has become common in the evaluation community.

This paper is organized in the following way. Section 2 defines the problem of selection bias in conducting nonexperimental evaluations of program impact. Section 3 discusses the nonexperimental estimators used in our re-analysis of the NSW data. We are limited by the fact that we only have access to grouped, rather than individual level, data and so a variety of available nonlinear nonexperimental estimators could not be employed. Nonetheless, the estimators we consider are representative of methods that have previously been used to evaluate training programs. Section 4 outlines procedures for testing the appropriateness of alternative nonexperimental evaluation methods. Section 5 reports the empirical performance of the proposed testing strategy. Section 6 presents concluding remarks.

2. THE PROBLEM OF SELECTION BIAS

Selection bias arises in evaluating the impact of training on mean earnings when the mean earnings of trainees would differ from the mean earnings of comparison group members even in the absence of training. Letting Y^*_{it} be earnings of individual i in period t in the absence of training and letting $d_i = 1$ if a person receives training and $= 0$ otherwise, selection bias is present if

$$E(Y^*_{it}|d_i=1) \neq E(Y^*_{it}|d_i=0),$$

where we assume that comparison group members would not be trainees ($d_i = 0$).

Let Y_{it} be the observed value of earnings for individual i at time t. Let α_{it} be the impact of training on person i at time t. We adopt the convention that training occurs in period k. Then

$$Y_{it} = Y^*_{it} + d_i\alpha_{it}, \quad t > k, \tag{2.1}$$

$$Y_{it} = Y^*_{it}, \quad t < k.$$

We focus on estimating the mean impact of training on the trained, i.e.,

$$E(\alpha_{it}|d_i=1) = E(Y_{it} - Y^*_{it}|d_i=1).$$

The special case of identical training impact for all persons assumes $\alpha_t = \alpha_{it}$ for all i so

$$\alpha_t = E(\alpha_{it}|d_i=1) = E(Y_{it} - Y^*_{it}|d_i=1).$$

This case is the focus of most of the attention in the literature.

The mean post-program earnings of trainees is

$$E(Y_{it}|d_i=1) = E(\alpha_{it}|d_i=1) + E(Y^*_{it}|d_i=1).$$

The mean post-program earnings of nontrainees is

$$E(Y_{it}|d_i=0) = E(Y^*_{it}|d_i=0).$$

The difference in mean earnings between trainees and non-trainees is

$$E(Y_{it}|d_i=1) - E(Y_{it}|d_i=0) = E(\alpha_{it}|d_i=1) + \left\{E(Y^*_{it}|d_i=1) - E(Y^*_{it}|d_i=0)\right\}.$$

The term in braces is the selection bias term. In the case of random assignment of persons to treatment

$$E(Y^*_{it}|d_i=1) = E(Y^*_{it}|d_i=0) = E(Y^*_{it}),$$

so the term in braces is zero.

When an experimental control group is not available, evaluations are conducted using various nonexperimental comparison groups. Previous analysts have used as comparison groups: (a) individuals who applied to the program and were rejected, (b) individuals who did not apply to the program, or (c) samples of persons "similar" to trainees from population samples such as the Current Population Survey (CPS) or Panel Survey of Income Dynamics (PSID). In samples of type (c), training status is typically unknown so that the comparison group may include some persons who are actually trainees. (This is the problem of "contamination bias"). Samples of type (a) and (b) are constructed conditional on $d_i = 0$. A variety of procedures have been proposed for eliminating the effect of selection bias on estimates of program impact.

3. ALTERNATIVE NONEXPERIMENTAL ESTIMATORS FOR MEASURING THE IMPACT OF TRAINING ON EARNINGS IN THE PRESENCE OF NONRANDOM ASSIGNMENT

To conform with the training literature, we confine our review of the available models to linear specifications of earnings equations. Suppose that Y^*_{it} is a linear function of a set of observed characteristics X_{it}, and unobserved characteristics represented by U_{it}. Thus

$$Y^*_{it} = X_{it}\beta + U_{it}, \tag{3.1}$$

where β is a vector of parameters. To simplify the exposition in this section and to conform with most of the literature, we assume that the training effect is invariant across individuals but not across time, so that $\alpha_{it} = \alpha_t$. We let $X_i = (X_{i1},\ldots,X_{iT})$, where T is the number of periods of data on X available for each observation. (Below, we also consider the case where the training effect is allowed to depend on regressors.) With these assumptions, the equation for observed earnings may be written (using 2.1) as

$$Y_{it} = X_{it}\beta + d_i\alpha_t + U_{it}, \text{ for } t = 0,\ldots,T. \tag{3.2}$$

We assume that $E(U_{it}|X_i) = 0$ for all i, t.

When assignment to training is nonrandom, selection bias in the estimation of α_t can arise because of dependence between d_i and U_{it}. That is, in a model without regressors,

$$E(U_{it}|d_i) \neq 0$$

which is equivalent to $E(Y^*_{it}|d_i=1) - E(Y^*_{it}|d_i=0) \neq 0$. In a model with regressors, selection bias is present if

$$E(U_{it}|d_i,X_i) \neq 0 \tag{3.3}$$

so

$$E(Y_{it}|d_i,X_i) \neq X_{it}\beta + d_i\alpha_t.$$

In this case, an ordinary least squares regression of Y_{it} on X_{it} and d_i will not yield consistent estimates of α_t (or β).

A stochastic relationship between d_i and U_{it} can arise for a variety of reasons. In the absence of random assignment, participation in a training program may be the result of decisions made by individuals eligible for the program, by the program administrators, or both. Whatever the decision making procedure, it can be described in terms of an index function framework. Let the index, IN_i, be a function of both observed (Z_i) and unobserved (V_i) variables. (The vector Z_i may include all of the variables in X_i.) For simplicity, we follow standard practice and assume that this function is linear in X_i and V_i:

$$IN_i = Z_i\gamma + V_i. \tag{3.4}$$

Then the i^{th} individual's training status is

$$d_i = \begin{cases} 1 \text{ if and only if } IN_i > 0, \\ \\ 0 \text{ otherwise.} \end{cases} \tag{3.5}$$

For simplicity, and without loss of essential generality, V_i is assumed to be independently and identically distributed across persons, where the distribution function of V_i is denoted as $F(v_i) = Pr(V_i < v_i)$. Assuming that V_i is distributed independently of Z_i, we may write $Pr(d_i=1|Z_i) = E(d_i|Z_i) = 1 - F(-Z_i\gamma)$ which Rosenbaum and Rubin (1983) call the "propensity score."

Alternative nonexperimental selection estimators of α_t augment the earnings function and selection rule given in equations (3.2), (3.4) and (3.5) with additional assumptions in order to undo the dependence between U_{it} and d_i (see (3.3)). Nonexperimental estimators differ in the assumptions imposed, the data required to implement such estimators, and their robustness to alternative sampling plans and measurement error. Many widely used selection bias estimators impose more assumptions than are required to recover α_t. When excess identifying conditions are invoked, they can be tested.

Heckman and Robb (1985, 1986) present a comprehensive summary of selection bias estimators which can be implemented in alternative types of data (e.g., cross section, repeated cross section and longitudinal data) and they consider the robustness of each estimator to alternative sampling plans and errors in measuring training status. Our empirical analysis of alternative nonexperimental estimators for measuring the impact of the National Supported Work Demonstration on earnings considers only a subset of the available estimators which can be implemented on the grouped data available to us.

Dependence between U_{it} and d_i can arise for one of two not necessarily mutually exclusive reasons: (a) dependence between Z_i and U_{it} or (b) dependence between V_i and U_{it}. We refer to the first case as <u>selection on observables</u> and the second case as <u>selection on unobservables</u>. The source of selection bias for any particular problem depends on the actual process used to select individuals. We consider each case in turn. Throughout we assume that different i subscripted random variables are independently distributed.

3.1 Selection on Observables

Selection on observables occurs when the dependence between U_{it} and d_i is due to a set of observed variables, Z_i, which influence selection into the program. Put more formally

$$E(U_{it}|d_i,X_i) \neq 0 \quad \text{and} \quad E(U_{it}|d_i,X_i,Z_i) \neq 0 \tag{3.6}$$

but

$$E(U_{it}|d_i,X_i,Z_i) = E(U_{it}|X_i,Z_i).$$

In this case, controlling for the observed selection variables (Z_i) solves the selection bias problem, i.e., removes the dependence between latent earnings and the training dummy, d_i. In particular, one can form estimators by noting that the appropriate conditional expectation function for earnings, conditioned on d_i, X_i, and Z_i is given by

$$E(Y_{it}|d_i,X_i,Z_i) = X_{it}\beta + d_i\alpha_t + E(U_{it}|d_i,X_i,Z_i) \tag{3.7}$$

$$= X_{it}\beta + d_i\alpha_t + E(U_{it}|X_i,Z_i).$$

Assuming knowledge of the functional form of $E(U_{it}|X_{it},Z_i)$, this term can be inserted in (3.2) and the resulting equation can be estimated by regression methods to obtain consistent estimates of α_t. Such estimators are members of the class of <u>control function estimators</u> proposed by Heckman and Robb (1985), where $E(U_{it}|X_i,Z_i)$ are called <u>control functions</u>. Nonparametric matching procedures which contrast the earnings of trainees and comparison group members based on the X_i and Z_i characteristics are predicated on assumption (3.6).

In our empirical analysis, we present estimates of NSW program impact using <u>linear control function estimators</u>. Linear control function estimators, first proposed by Barnow, Cain and Goldberger (1980, pp. 47-48), are a natural starting point and, if (3.7) holds along with the linearity of $E(U_{it}|X_i,Z_i)$, produce consistent estimators of program impacts. We use two variants of their estimator. <u>Variant I</u> corresponds to the assumption that $E(\alpha_{it}|d_i=1,X_i,Z_i) = \alpha_t$. In this case, inserting a linear version $E(U_{it}|X_i,Z_i)$ in (3.2) yields:

$$Y_{it} = C_i\delta_t + d_i\alpha_t + \tilde{U}_{it} \tag{3.8}$$

where C_i denotes the vector of all variables included in either X_i or Z_i, $\tilde{U}_{it} = U_{it} - E(U_{it}|d_i,C_i) = U_{it} - E(U_{it}|C_i)$ and δ_t is a parameter vector. Consistent estimators of α_t can be obtained by using either ordinary or weighted [by the square root of cell size] least squares to estimate (3.8). <u>Variant II</u> of this estimator allows the training impact to depend on person-specific values of the regressors, i.e, $E(\alpha_{it}|d_i=1) = C_i\theta_t$. The appropriate estimating equation is

$$Y_{it} = C_i\delta_t + d_i(C_i\theta_t) + \tilde{U}_{it}. \tag{3.8'}$$

3.2 Selection on Unobservables

It may happen that the dependence between the training indicator variable, d_i, and U_{it} is not eliminated even after controlling for Z_i. That is:

$$E(U_{it}|d_i,X_i) \neq 0 \text{ and } E(U_{it}|d_i,X_i,Z_i) \neq E(U_{it}|X_i,Z_i). \tag{3.9}$$

Selection is then said to depend on unobservables. A number of estimation procedures have been proposed to deal with selection bias when selection is on unobservables. Such estimators are formed by invoking assumptions about the distributions, or moments of the distributions, of V_i, Z_i and U_{it}. In our analysis of the NSW data, we consider two estimators, the fixed effect and random growth estimators. Each is appropriate under a specific set of assumptions about the form of the dependence between V_i and U_{it}.

First consider the <u>fixed effect</u> or first difference estimator. Suppose that, even though (3.9) holds, the conditional expectation of the difference in a pre- and post-training set of U_{it}'s does not depend on d_i. That is, assume that the following condition holds:

$$E(U_{it} - U_{it'}|d_i,X_i) = 0, \text{ for all } t,\ t',\ t > k > t'. \tag{3.10}$$

This specification is motivated by a model in which U_{it} is of the form

$$U_{it} = \phi_{1i} + \nu_{it},$$

where ϕ_{1i} is a zero mean, person-specific component or "fixed effect," ν_{it} is a zero mean random component independent of all other values of $\nu_{it'}$ $(t \neq t')$

and of ϕ_{1i}. In this specification, selection is assumed to occur on the permanent component ϕ_{1i}, i.e., this component accounts for the dependence between V_i and U_{it}.

For this model, consistent estimates of the impact of training can be obtained by regressing the difference between Y_{it} and $Y_{it'}$ on d_i and $X_{it} - X_{it'}$. As we do with the linear control function estimators, we estimate two variants of the fixed effect estimator. Variant I of the fixed effect estimator is derived by estimating the following regression

$$Y_{it} - Y_{it'} = d_i\alpha_t + (X_{it} - X_{it'})\beta + (\nu_{it} - \nu_{it'}), \text{ for } t > k > t'. \quad (3.11)$$

If (3.10) holds, under standard rank conditions this produces a consistent estimator of α_t. Variant II, which again allows the training impact to vary as a function of C_{it}, is obtained by estimating the following regression:

$$Y_{it} - Y_{it'} = d_i(C_i\theta_t) + (X_{it} - X_{it'})\beta + (\nu_{it} - \nu_{it'}), \text{ for } t > k > t'. \quad (3.11')$$

We also use a random growth estimator, which is motivated by a special case of a more general class of models in which U_{it} has a factor structure. In particular, suppose U_{it} is of the following form:

$$U_{it} = \phi_{1i} + t\phi_{2i} + \nu_{it}, \quad (3.12)$$

where ϕ_{1i} is as before and ϕ_{2i} is a person-specific growth rate where (ϕ_{1i}, ϕ_{2i}) are assumed to have zero means and finite variances and to be independent of ν_{it} for all i and t. The dependence between U_{it} and d_i is assumed to arise because of dependence between d_i and (ϕ_{1i}, ϕ_{2i}).

Given (3.12), one can transform the earnings equation to eliminate ϕ_{1i} and ϕ_{2i} from the earnings equation. Values of Y_{it} in two consecutive pre-training periods can be used to proxy these components. We obtain Variant I of the random growth estimator by estimating the following equation

$$(Y_{it} - Y_{it'}) - (t-t')(Y_{it'} - Y_{i,t'-1}) = d_i\alpha_t + [(X_{it} - X_{it'}) -$$

$$(t-t')(X_{it'} - X_{i,t'-1})]\beta + [(\nu_{it} - \nu_{it'}) - (t-t')(\nu_{it} - \nu_{i,t'-1})], \quad (3.13)$$

for $t > k > t'$, by least squares. The resulting estimator of α_t is consistent under standard conditions. (Pudney (1982) proves that the asymptotic distribution of the estimator of α_t is invariant to the choice of earnings from other periods used to proxy for ϕ_{1i} and ϕ_{2i} provided that all the ν_{it}'s have non-zero variances and the equation is estimated by generalized least squares.) As before, <u>Variant II</u> of the random growth estimator is obtained by estimating the following regression

$$(Y_{it} - Y_{it'}) - (t-t')(Y_{it'} - Y_{i,t'-1}) = d_i(C_i\theta_t) + [(X_{it} - X_{it'}) -$$

$$(t-t')(X_{it'} - X_{i,t'-1})]\beta + [(\nu_{it} - \nu_{it'}) - (t-t')(\nu_{it} - \nu_{i,t'-1})], \quad (3.13')$$

for $t > k > t'$.

The fixed effect, random growth and linear control function estimators all yield consistent estimators of the training effect when applied to choice-based samples because they are based on conditional (on d_i) moment restrictions. Choice-based samples are samples that mix trainees and comparison group members in different proportions than they are found in the population (see, e.g., Manski and McFadden, 1981). Our sample is choice-based so the robustness of estimators to this sampling plan is a desirable feature.

4. TESTING ALTERNATIVE SPECIFICATIONS

We propose and implement three types of model specification tests. The first is premised on access to data on pre-program earnings and regressor variables for future program participants (trainees and, when available, controls) and comparison group members. Such data are widely available. Ignoring contamination bias, a candidate selection correction procedure for program evaluation applied to pre-program data should make the adjusted earnings equation of future trainees and comparison group members alike provided that the equation for pre-program earnings is like that for post-program earnings except for the additive training effect. If a candidate selection bias adjustment does not align the pre-program earnings equations for future participants (trainees and controls) and comparison group members and if it is plausible to assume that the source of pre-program differences in earnings between the two types of individuals is the same as for the post-program differences, the candidate correction procedure is rejected.

Our second test is based on overidentifying restrictions implied by certain models. Even in the absence of pre-program data on earnings for participants and comparison group members, it is sometimes possible to test the validity of a particular selection adjustment procedure. Heckman and Robb (1985) note that the assumption of normality or symmetry for U_{it} which underlies many adjustment procedures can be tested in a single post-program cross section of program trainees and comparison group members. Many other selection estimators are based on assumptions which can be subject to test. Rejection of the testable assumptions underlying a procedure would cause rejection of a candidate selection correction method.

A third test of the validity of a nonexperimental estimator is premised on access to experimental data. Controls from the experiment, who do not receive training, are pooled with comparison group members. Controls are like trainees except they do not receive training. Ignoring contamination bias, a valid selection correction procedure should make the adjusted earnings equations for controls and comparison group members alike. Unlike the first test, this test does not assume temporal stability in the earnings equation and for this reason is more robust than a test based on pre-program earnings.

The third test is of no direct use in any particular nonexperimental evaluation because, by assumption, an experimental estimate is available. Its value comes in evaluating an estimator that might be suitable for nonexperimental evaluation of the same program when experimental data are not available or in picking an estimator for a similar program.

4.1 The Pre-Program Tests

The linear control function version of this test is based on equations (3.8) and (3.8′), where $t < k$ and where $d_i = 1$ if an observation is a future participant (trainee or control) and $d_i = 0$ if an observation is from the comparison group. If a valid linear control function has been used, estimated values of α_t and θ_t $(t < k)$ should not be statistically significantly different from zero since no observation has undertaken training.

The fixed effect ((3.11) and (3.11′)) and random growth ((3.13) and (3.13′)) versions of this test modify these equations and use pre-program information for period t and t'. Defining d_i as in the immediately preceding paragraph, estimated values of α_t and θ_t should not be statistically significantly different from zero for any correctly specified selection correction model.

4.2 The Post-Program Tests

These tests are identical in structure to the pre-program tests except now $t > k$ and $d_i = 1$ if an observation is a member of the experimental control group and $d_i = 0$ if an observation is a member of the comparison group. Since neither group receives training, estimated values of α_t and θ_t should not be statistically significantly different from zero for a valid nonexperimental selection correction estimator.

4.3 Tests of Overidentifying Restrictions

In the absence of strong beliefs about the functional form of the outcome equation and the appropriate regressor variables, there are no testable restrictions (apart from those already discussed) implied by the linear control function estimator. The situation is different for the fixed effect and random growth estimators. Under the assumptions that justify those estimators, values of Y from periods other than those specified by equations (3.8), (3.8′), (3.13) and (3.13′) should not appear as regressors in those equations. A test that the coefficients on these extraneous Y values are equal to zero is a test of the overidentifying restrictions implied by these models. For these models to be overidentified, there must be enough periods of panel data for there to be extraneous Y variables. Thus, in samples with two periods of panel data, the fixed effect estimator is just-identified and has no testable restrictions apart from those already presented.

5. A RE-ANALYSIS OF THE NATIONAL SUPPORTED WORK DATA

In this section, we apply our specification tests to models estimated on data from the National Supported Work (NSW) experiment previously analyzed by LaLonde (1986), Fraker and Maynard (1984, 1987), and LaLonde and Maynard (1987) in their critiques of nonexperimental evaluation procedures. All NSW participants (both trainees and controls) used in our analysis were enrolled in the program in either 1976 or 1977. Fraker and Maynard obtain both pre- and post-training data on earnings from Social Security Administration (SSA) records for the NSW participants and comparison groups drawn from the March 1976 and March 1977 Current Population Survey (CPS). The comparison group is temporally aligned with the NSW participant group, so that the "pre" and "post" program periods for this group refer to the time periods corresponding to the period of operation of the NSW program.

To protect the confidentiality of the SSA earnings information on individuals, only mean values of earnings for cells of individuals for both the NSW participants and those in the CPS were provided by the Social Security Administration. The means are for cells consisting of 7 to 10 sample members from the NSW trainees and controls and the CPS comparison groups, respectively, where sample members were assigned to cells on the basis of: (a) date of enrollment (for NSW participants) or date of interview (for CPS comparison groups), (b) whether or not they had been employed in the year prior to enrollment or interview and (c) geographical location. We used the grouped version of these data because they provide extensive longitudinal earnings histories. The price of using these histories is that the grouping of the data precludes the use of many nonlinear nonexperimental estimators. Many control functions are nonlinear functions of X_i and Z_i. Such estimators will generally not be consistent when using grouped rather than individual level data. The mean of a nonlinear function of X's is not equal to the nonlinear function evaluated at their mean values. Estimators which are linear functions of regressors will be consistent using either grouped or individual level data. Therefore, we restrict our investigation to linear nonexperimental estimators.

For high school dropouts, we use the comparison group constructed by Fraker and Maynard from the CPS that consists of individuals who (a) were between age 16 and 20 in the interview or enrollment year, (b) were not in school in the interview month or at enrollment, and (c) were high school dropouts. For AFDC Females, we also use a CPS-based comparison group consisting of adult women who (a) are between age 18 and 64 in the interview or enrollment year, (b) are AFDC recipients and (c) have dependents age sixteen or less. The criteria used by Fraker and Maynard to define both comparison groups mimic the eligibility requirements for the program. Given the highly geographically concentrated nature of NSW and the small fraction of the U.S. population living in the NSW target areas, it is plausible to assume that few, if any, persons in either comparison group were eligible to participate in the NSW program so the problem of contamination bias is unimportant in our data.

The variables used in our analysis are defined in Table 1. Mean values for trainees, controls and the comparison group for both Youths and AFDC women are given in Table 2. Table 2 demonstrates how random assignment of both Youth and AFDC Female NSW participants produces trainees and controls with virtually identical pre-training characteristics. In contrast, the means of most of the variables differ substantially between the CPS comparison group and the NSW participant groups (trainees and controls) for Youth. For AFDC females there is closer agreement between the means of the CPS comparison group and the means of the NSW participants but the discrepancies are still sizeable.

5.1 Estimates of the Impact of Training

Tables 3 and 4 present estimates of program impact for high school dropouts and AFDC women, respectively, produced from the two variants of each of the three basic models presented in Section 3. The format is the same for each table. For both variants of each model, we present estimates (and their standard errors) of the impact of training on post-training earnings in 1978 and 1979. For each year, the columns labelled Variant I record estimates of α_t and the columns labelled Variant II present estimates of the model in which training impact is of the form $\alpha_{it} = C_i\theta_t$ and is evaluated at the sample means for C_i from the NSW trainee sample ($\bar{C}$). The rows in these tables give the training impact estimates using:

the linear control function estimator (equations (3.8) and (3.8′));

the fixed effect estimator (equations (3.11) and (3.11′)) constructed using both 1972 earnings and 1974 earnings for the pre-program earnings to construct the pre- and post-program differences;

the random growth estimator (equations (3.13) and (3.13′)) constructed using using both 1972-73 earnings and 1973-74 earnings for the pre-program earnings to construct the dependent variable for this estimator;

and

the experimental estimator (formed by the difference in weighted means of NSW trainees and NSW controls, where the weights used to compute these means are the square roots of the number of individuals in the grouped data cells).

Estimates are presented for each nonexperimental estimator controlling for alternative sets of X_i (and C_i), which are described in Table 1 and labeled B1, B2, W1, W2. The notation B1 + B2 denotes the combined set of variables from B1 and B2. All estimates are obtained using $\sqrt{N_j}$-weighted least squares to account for the grouped nature of the data where N_j is the number of observations in cell j. The standard errors are produced from conventional formulae. The same inferences are obtained from jackknifed standard errors which are not reported here.

The bottom row in both Tables 3 and 4 presents the experimental estimates for high school dropout and AFDC women, respectively. For high school dropouts, the weighted mean experimental differences are -48 in 1978 and 9 for 1979. Neither estimate is statistically significantly different from zero. For AFDC women, there is evidence of a statistically significantly positive impact of training on 1978 earnings (440) and a weaker positive effect on 1979 earnings (267). The weighted mean difference between trainees and the nonexperimental comparison group are found in the first row of the linear control function estimates labelled "No Control Variables" in the two tables. The estimates for high school dropouts are statistically significantly negative for each year (-1910 and -1917, for 1978 and 1979, respectively), while for AFDC women, they are small and statistically insignificant (157 and 79, for 1978 and 1979, respectively).

The nonexperimental estimates presented in these two tables exhibit the same kind of instability chronicled by LaLonde (1986). Different nonexperimental estimators produce very different inferences about the effect of the program. For Youth, the nonexperimental estimates are always negative--often statistically significantly so. For AFDC women, the nonexperimental estimates are generally positive and often statistically significantly so, especially for the 1978 earnings measure.

Tables 3 and 4 suggest that selection bias is an empirically important problem in using nonexperimental data to evaluate the impact of training on earnings. If selection bias were not present, alternative nonexperimental methods would generate the same inference about the impact of training. The fact that alternative nonexperimental estimators produce different inferences about training indicates that some--perhaps all--of the models are misspecified. To see if it is possible to detect misspecified models, we now turn to the results from our specification tests. We consider results for Youth and Women in turn. We note that a limited set of the pre-program tests presented below are also found in Heckman, Hotz and Dabos (1987).

5.2 Results of Model Selection Tests for High School Dropouts (Youth)

Table 5 reports probability values (P values) for the specification tests described in Section 4. Under the heading "Pre-Program Tests Using Pre-Program Earnings," we present P values for the hypotheses $\alpha_t = 0$ (Variant I) and $\bar{C}\theta_t = 0$ (Variant II), $t < k$, for earnings models fit on a pooled sample of future participants (trainees and controls) and comparison group members. Recall that for this test, $d_i = 1$ if an observation is a trainee (or control) and $= 0$ otherwise. Tests for the vector hypotheses $\theta_t = 0$ are always consistent with tests for the more restricted hypothesis $\bar{C}\theta_t = 0$ and for the sake of brevity are not reported.

Under the heading "Post-Program Tests Using ...," we present P values for tests of the hypotheses $\alpha_t = 0$ and $\bar{C}\theta_t = 0$, $t > k$, for earnings models fit on a pooled sample of experimental controls from the experiment ($d_i = 1$) and comparison group members ($d_i = 0$). Again, tests of the vector hypothesis $\theta_t = 0$ are consistent with the test based on $\bar{C}\theta_t$ and are not reported here.

Under the heading "Overidentification Tests Using ...," we report P values for the hypotheses that extraneous Y values do not have statistically significant coefficients in the fixed effect and random growth models fit on pre-program earnings (for a pooled sample of future trainees and controls ($d_i = 1$) and comparison group members ($d_i = 0$)) and on post-program earnings (for a pooled sample of controls from the experiment ($d_i = 1$) and comparison group members ($d_i = 0$)). We do not see any compelling overidentifying restriction for the linear control function estimator and so report no test of overidentification for that model.

The pre- and post-program tests and the overidentification tests generally produce consistent findings. Linear control function and fixed effect models are decisively rejected. The random growth model is not. (While not reported in Table 5, the coefficient estimates associated with the pre-program and post-program tests range from -2128 to -167 for the linear control function models, from -2186 to -274 for the fixed effect models, and from -894 to -37 for the random growth models.) However, the tests for the overidentifying restrictions applied to the random growth model fit on post-program data are mixed. Using 1973 and 1972 earnings to proxy the unobserved components, ϕ_{1i} and ϕ_{2i}, the model is rejected on the post-program data. Using 1973 and 1974 earnings to proxy the unobserved components, the model is not rejected on the post-program data. Neither version of the random growth model with regressors is rejected when it is fit on the pre-program sample which combines future participants (trainees and controls) and comparison group members.

The rejected version of the model uses the longest lags in pre-program earnings of any of the fitted models to eliminate the permanent and random growth components in U_{it} ($t-t' = 6$ and 7 for 1978 and 1979, respectively). Earnings functions are well known to be concave in age or experience. Linear growth specification (3.12) may become a progressively poorer approximation as the lag length increases between the dependent variable and the proxy variables. A better model might augment (3.12) to include a third component, ϕ_{3i}, multiplied by $(t - t')^2$. To find sufficient proxy variables for this model requires a third year of pre-program earnings data which is not available to us. Models with autoregressive specifications for U_{it} failed specification tests.

A slight extension of (3.12) produces a model that passes specification tests and produces estimates of program impact very close to those obtained from the random growth model reported in Table 3. In place of (3.12), we write

$$U_{it} = \phi_{1i} + b_t\phi_{2i} + \nu_{it}, \; b_t \neq b_{t'}, \; t \neq t',$$

which permits the growth component to be unrestricted. The coefficient on $Y_{it'} - Y_{i,t'-1}$ in (3.13) and (3.13′), defined to be $\omega_{t',t'-1} = (b_t - b_{t'})/(b_{t'} - b_{t'-1})$, now becomes a parameter to be estimated. Moving $(Y_{it'} - Y_{i,t'-1})$ multiplied by this coefficient to the right side of those equations and utilizing instrumental variables to account for endogeneity of the regressor produces a model that is not rejected by any of the post-program specification tests. (See Table 6.)

The endogeneity in this variable is due to its dependence on $\nu_{it'}$ and $\nu_{i,t'-1}$ which appear as disturbances in the equations. Variables in B2 are used as instruments assuming that the variables in B1 belong in the earnings equation. Empirical results based on these instruments must be qualified in light of the finding by Heckman and Robb (1985) that unweighted instrumental variables procedures are not robust to choice-based sampling. We do not have access to the data required to construct the appropriate weights to guarantee consistency of the instrumental variables estimator.

The estimates of the free parameters on $(Y_{it'} - Y_{i,t'-1})$, reported as "$\omega_{t',t'-1}$" in Table 7, are not exactly equal to (t-t′), but one cannot reject the hypothesis that they are equal to (t-t′) at conventional levels of significance. Reading across the first row of Table 7, the estimated values of $\omega_{t',t'-1}$ for 1978 earnings are close to $(t-t') = 6$ while the estimated values of $\omega_{t',t'-1}$ for 1979 earnings are close to $(t-t') = 7$. The true specification for Youth in this program seems to be quite close to the random growth model. The estimated program impacts in this table are close to those reported in Table 3 using B1 variables as regressors.

Returning to Table 3, note that the random growth model produces the same inference about program impact as the experiment--that training has no effect on earnings. The estimates from the random growth model are more negative than the estimates produced by the experiment but the standard errors for the nonexperimental estimator are also bigger.

Since we have no basis for distinguishing among the various versions of the random growth model which control for alternative sets of variables in X_i and survive our battery of tests, we compute a "Weighted Average of Estimates" for the random growth models fit using alternative sets of pre-program years for t' and $t'-1$. The weighted average is formed by applying generalized least squares to the regression coefficients of each model, treated as observations, using the variance-covariance matrix of each regression coefficient to form the variance for each observation. Such weighted averages can be given a Bayesian justification. (See Leamer, 1978.) The weighted averages, based on the alternative random growth estimators which are not rejected by the post-program and over-identification tests using 1978 and 1979 earnings, are presented at the bottom of Table 3. Although the point estimates differ from those obtained from experimental data, they produce the same conclusion as is obtained from the experiment, namely that there is no statistically significant impact of the NSW program on Youth earnings. The estimates from the modified random growth estimator presented in Table 7 also produce the same inference and are closer to the experimental results. (Also note that exactly the same random growth estimators are rejected by the pre-program tests and tests of overidentifying restrictions based on pre-program earnings data. Therefore, even in the absence of the experimental data, one would have chosen the same set of random growth estimators and thus obtained the same weighted average estimates as in Table 3.)

Observe that the higher the P value for an estimator displayed in Tables 5 and 6, the lower, on average, is the discrepancy between the nonexperimental and experimental estimates. Low P values indicate model misspecification. Such misspecification should widen the difference between the estimate obtained from the misspecified model and the experimental estimate.

For Youth our testing strategies lead to a quite different conclusion than that reported by Fraker and Maynard and LaLonde. The models that survive our tests yield the same conclusion about the impact of the NSW training program on the earnings of Youths as is obtained from the experiment. The rejected models are the source of the discrepancy in inference between experimental and nonexperimental estimates reported in the literature. Note, however, that the nonexperimental estimators not rejected by these specification tests have much larger standard errors than the experimental estimators. Experimental data produce a sharper inference. Nonetheless, our results for Youth suggest that pessimism concerning the use of nonexperimental estimators may not be well founded and that a systematic procedure exists to identify estimators that replicate the inferences drawn from experimental methods.

5.3 Results of Model Selection Tests for AFDC Recipients (Women)

Table 8 reports the results of specification tests applied to alternative earnings equations for AFDC women. The format of this table is the same as that of Table 5. Neither the pre-program tests nor the post-program model specification tests are decisive in rejecting any of the models. (The coefficient estimates associated with these tests range from -686 to 476 for the linear control function models, -449 to 750 for the fixed effect models and -1961 to 1071 for the random growth models.) The tests of overidentifying restrictions have much more bite. For both the pre-program and post-program versions of these tests, the fixed effect and random growth models are rather decisively rejected. By default, we do not reject the linear control function estimators.

Returning to Table 4, we examine the inference about the impact of the experiment which is derived from the linear control function specification. A generalized least squares average of the linear control function estimators which are not rejected by the post-program tests produces the number shown in the row labelled "Weighted Average of Estimates." (In the absence of experimental data, a slightly different set of linear control function estimators not rejected by the pre-program tests would be included in this average. In this case, the corresponding average estimates are 702 with a standard error of 168 for 1978 earnings and 515 with a standard error of 179 for 1979 earnings.) The 1978 weighted nonexperimental estimate, while somewhat lower than the experimental estimate, leads to the same strong inference--that training raised the earnings of trainees. The 1979 weighted nonexperimental estimate also leads to an inference similar to that obtained from the experiment--that of a slightly weaker, but still positive, effect of training on earnings. Again, the source of the difference in inference between experimental and nonexperimental estimates arises from the rejected models.

6. CONCLUSIONS

This paper considers the problem of assessing the validity of alternative nonexperimental evaluation estimators. We critically examine the claims of LaLonde and Fraker and Maynard concerning the difficulty in using nonexperimental methods to evaluate social programs. A simple model selection strategy based on easily implemented specification tests eliminates nonexperimental evaluation models that do not produce estimated program impacts close to the experimental results: the models that are not rejected produce impacts that are close to the experimental results, at least in the case of women on AFDC.

We do not claim that we have found the "true" model for either Youth or AFDC Women. Such a claim would be premature. As noted above, a variety of nonlinear nonexperimental estimators could not be implemented in this study because of the grouped nature of the data available to us. Nonetheless, using the same data set analyzed by critics of nonexperimental data, our analysis demonstrates that simple specification tests eliminate the most unreliable and misleading estimators that give rise to the "sensitivity" problem recently discussed in the evaluation literature. Thus, while not definitive, our results are certainly encouraging for using nonexperimental methods for social program evaluation.

REFERENCES

Ashenfelter, O. and Card, D. (1985), "Using the Longitudinal Structure of Earnings to Estimate the Effect of Training Programs." Review of Economics and Statistics, 67, 648-660.

Barnow, B. (1987), "The Impact of CETA Programs on Earnings: A Review of the Literature." Journal of Human Resources, XXII, 157-193.

Barnow, B., Cain, G., and Goldberger, A. (1980), "Issues in the Analysis of Selectivity Bias." Evaluation Studies, eds. E. Stromsdorfer and G. Farkas, 1980, 5, 42-59.

Barros, R. (1987), Two Essays on the Nonparametric Estimation of Economic Models with Selectivity Using Choice-Based Samples, Unpublished Ph.D. Dissertation, University of Chicago.

Burtless, G. and Orr, L. (1986), "Are Classical Experiments Needed for Manpower Policy?" Journal of Human Resources, 21, 606-639.

Fraker, T. and Maynard, R. (1984), An Assessment of Alternative Comparison Group Methodologies for Evaluating Employment and Training Programs. Princeton: MPR, Inc.

Fraker, T. and Maynard, R. (1987), "Evaluating Comparison Group Designs with Employment-Related Programs." Journal of Human Resources, 22, 194-227.

Heckman, J., Hotz, J. and Dabos, M. (1987), "Do We Need Experimental Data to Evaluate the Impact of Manpower Training on Earnings?" Evaluation Review, 11, 395-427.

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

Heckman, J. and Robb, R. (1986), "Alternative Identifying Assumptions in Econometric Models of Selection Bias." in Advances in Econometrics: Innovations in Quantitative Economics. Essays in Honor of Robert L. Basmann, 5, eds. D. Slottje, Greenwich, CT: JAI Press, 243-287.

LaLonde, R. (1986), "Evaluating the Econometric Evaluations of Training Programs with Experimental Data." American Economic Review, 76, 604-620.

LaLonde, R. and Maynard, R. (1987), "How Precise are Evaluations of Employment and Training Programs: Evidence from a Field Experiment." Evaluation Studies, 11, 428-51.

Leamer, E. (1978), Specification Searches: Ad Hoc Inference with Nonexperimental Data, New York: John Wiley.

Manski, C. and McFadden, D. (1981), "Alternative Estimators and Sample Designs for Discrete Choice Analysis," in Structural Analysis of Discrete Data with Econometric Applications, eds. C. Manski and D. McFadden, Cambridge, MA: MIT Press, 2-50.

Pudney, S. (1982), "Estimating Latent Variable Systems When Specification is Uncertain: Generalized Component Analysis and the Eliminant Method." Journal of the American Statistical Association, 77, 883-889.

Rivlin, A. (1971), Systematic Thinking for Social Action. Washington, D.C.: The Brookings Institution.

Rosenbaum, P. and Rubin, D. (1983), "The Central Role of The Propensity Score in Observational Studies For Causal Effects." Biometrika, 70, 41-55.

Rossi, P. and Freeman, H. (1985), Evaluation: A Systematic Approach. Third Edition. Beverly Hills, CA: SAGE Publications, Inc.

Stromsdorfer, E., et al. (1985), "Recommendations of the Job Training Longitudinal Survey Research Advisory Panel," Report to the Employment and Training Administration, U.S. Department of Labor.

Table 1. Definition of Variables

Variable Name	Description
Earnings Variables:	
SSEARN72	SSA earnings in 1972 (in 1978 dollars)
SSEARN73	SSA earnings in 1973 (in 1978 dollars)
SSEARN74	SSA earnings in 1974 (in 1978 dollars)
SSEARN75	SSA earnings in 1975 (in 1978 dollars)
SSEARN78	SSA earnings in 1978 (in 1978 dollars)
SSEARN79	SSA earnings in 1979 (in 1978 dollars)
Background Variables in B1:	
BLKHIS	= 1 if black or hispanic and = 0 otherwise
SEX	= 1 for men and = 0 for females
MARRIAGE	= 1 if married at enrollment for NSW Participants or at March interview for CPS Respondents and = 0 otherwise
AGE	Age in years at enrollment for NSW Participants or at March interview for CPS Respondents
EDUC	Years of schooling completed at enrollment for NSW Participants or at March interview for CPS Respondents
URBAN	= 1 if in central city SMSA, = 0 otherwise
7677ENR	= 1 if enrolled in 1977 for NSW Participants or if interviewed in March 1977 for CPS Respondents and = 0 otherwise
Background Variables in B2:	
BLACK	= 1 if black and = 0, otherwise
HISPANIC	= 1 if hispanic and = 0, otherwise
AGESQ	AGE squared
HOUSESIZE	Number of household members at enrollment for NSW Participants or at March interview for CPS Respondents
DEPEND	Number of dependents at enrollment for NSW Participants or at March interview for CPS Respondents (used only for AFDC Recipient (Women) results)
AGEKID	Age of youngest dependent at enrollment for NSW Participants or at March interview for CPS Respondents (used only for AFDC Recipient (Women) results)
Work History Variables in W1:	
SSEARNL1	Annual earnings (from SSA data) one year prior to enrollment for NSW Participants or one year prior to interview of CPS Respondents
SSEARNL2	Annual earnings (from SSA data) two years prior to enrollment for NSW Participants or two years prior to the interview for CPS Respondents
WORKWKS	Number of weeks worked in year prior to enrollment for NSW Participants or in year prior to interview for CPS Respondents
UEWKS	Number of weeks unemployed in year prior to enrollment for NSW Participants or in year prior to interview for CPS Respondents
AVEHRS	Average hours per week (when worked) in year prior to enrollment for NSW Participants or in year prior to interview for CPS Respondents
WELFARE	Per capita benefit the household received from Welfare and earnings of the sample member in the month prior to enrollment for NSW Participants or the interview for CPS Respondents
Work History Variables in W2:	
SSEARNL3	Annual earnings (from SSA data) three years prior to enrollment for NSW Participants or three years prior to interview for CPS Respondents
SSEARNL4	Annual earnings (from SSA data) four years prior to enrollment for NSW Participants or four years prio to interview for CPS Respondents
CLERSALE	= 1 if job prior to enrollment for NSW Participants or if current/most recent job for CPS Respondents was a clerical or sales occupation and = 0 otherwise
SERVICE	= 1 if job prior to enrollment for NSW Participants or if current/most recent job for CPS Respondents was in service sector and = 0 otherwise
PROFESSION	= 1 if job prior to enrollment for NSW Participants or if current/most recent job for CPS Respondents was a professional occupation and = 0 otherwise
AFDC	= 1 if received AFDC in year prior to enrollment for NSW Participants or in year prior to interview fo CPS Respondents and = 0, otherwise (used only for High School Dropout (Youth) results)

Table 2. Sample Means

Variable Name	High School Dropouts (Youth) NSW Samples Trainees	Controls	CPS Sample	AFDC Recipients (Women) NSW Samples Trainees	Controls	CPS Sample
Earnings Variables:						
SSEARN72	192.7	228.9	201.3	971.3	1085.6	1041.0
SSEARN73	329.9	401.1	548.4	1087.6	1206.3	1192.7
SSEARN74	581.1	630.6	1036.6	895.3	1000.8	1201.9
SSEARN75	532.4	504.9	1455.9	541.4	638.9	1045.8
SSEARN78	1704.0	1751.5	3654.8	2007.8	1588.9	1841.8
SSEARN79	1838.2	1825.5	3787.0	2039.9	1798.3	1959.6
Background Variables in B1:						
BLKHIS	0.918	0.909	0.196	0.955	0.945	0.500
SEX	0.883	0.864	0.483	0.000	0.000	0.000
MARRIAGE	0.044	0.033	0.285	0.023	0.042	0.186
AGE	18.200	18.347	18.080	33.375	33.615	31.460
EDUC	9.616	9.677	10.658	10.307	10.272	11.133
URBAN	1.000	1.000	0.240	0.979	0.981	0.440
7677ENR	0.645	0.651	0.433	0.729	0.719	0.485
Background Variables in B2:						
BLACK	0.736	0.706	0.110	0.835	0.817	0.381
HISPANIC	0.182	0.203	0.086	0.120	0.128	0.120
HOUSESIZE	4.704	4.746	3.335	3.613	3.779	3.636
DEPEND				2.167	2.292	2.506
AGEKID				9.341	9.215	10.124
AGESQ				1169.06	1181.43	1078.44
Work History Variables in W1:						
SSEARNL1	559.0	539.3	1545.4	459.0	462.1	905.3
SSEARNL2	436.4	447.4	944.8	508.8	607.6	862.8
WORKWKS	9.3	9.3	21.4	3.3	3.2	11.0
UEWKS	10.3	11.2	2.2	11.7	13.3	1.9
HOURS	3.3	3.2	15.9	1.3	0.9	7.2
WELFARE	33.8	33.7	121.6	93.9	91.6	169.7
Work History Variables in W2:						
SSEARNL3	357.1	400.6	521.9	711.8	816.5	872.5
SSEARNL4	180.6	198.1	194.7	711.9	769.1	741.1
CLERSALE	0.101	0.108	0.126	0.072	0.081	0.109
SERVICE	0.256	0.212	0.228	0.100	0.084	0.164
PROFESSION	0.057	0.040	0.016	0.013	0.016	0.022
AFDC	0.045	0.043	0.146			
No. of Obs.	566	678	2368	800	802	1995
No. of Cells	69	87	321	110	107	266

Table 3. Estimates of Training Effects for High School Dropouts (Youths)

Model and Control Variable Sets	1978 Earnings Variant I (α_t)	1978 Earnings Variant II $(\dot{C}\theta_t)$	1979 Earnings Variant I (α_t)	1979 Earnings Variant II $(\dot{C}\theta_t)$
Nonexperimental Estimates:				
Linear Control Function Estimates:				
No Control Variables	-1910 (243)		-1917 (191)	
B1	-1884 (247)	-1827 (246)	-2119 (342)	-2092 (300)
B1+B2	-1279 (273)	-1079 (295)	-1569 (239)	-1498 (319)
B1+W1	-1117 (246)	-1146 (263)	-1539 (343)	-1447 (372)
B1+B2+W1+W2	-889 (328)	-889 (380)	-996 (442)	-1331 (388)
Fixed Effect Estimates Constructed with t'=1972 Pre-Training Earnings:				
No Control Variables	-1904 (236)		-1910 (266)	
B1	-1886 (242)	-1831 (201)	-2172 (277)	-2070 (275)
B1+B2	-1360 (270)	-1227 (291)	-1644 (309)	-1647 (330)
Fixed Effect Estimates Constructed with t'=1974 Pre-Training Earnings:				
No Control Variables	-1456 (203)		-1462 (166)	
B1	-1411 (227)	-1370 (228)	-1663 (301)	-1636 (269)
B1+B2	-1035 (255)	-964 (276)	-1330 (326)	-1383 (312)
Random Growth Estimates Constructed with t'=1973, t'-1=1972 Pre-Training Earnings:				
No Control Variables	-649 (336)		-446 (386)	
B1	-231 (414)	-235 (416)	-241 (475)	-236 (477)
B1+B2	-23 (476)	76 (515)	-85 (547)	-126 (589)
Weighted Average of Estimates	-24 (185)		-154 (212)	
Random Growth Estimates Constructed with t'=1974, t'-1=1973 Pre-Training Earnings:				
No Control Variables	-499 (328)		-267 (307)	
B1	-614 (431)	-589 (436)	-701 (510)	-659 (515)
B1+B2	-624 (497)	-777 (537)	-806 (586)	-850 (630)
Weighted Average of Estimates	-616 (426)		-724 (502)	
Experimental Estimates				
	-48 (144)		9 (173)	

NOTE: Standard Errors in Parentheses.

Table 4. Estimates of Training Effects for AFDC Recipients (Women)

Model and Control Variable Sets	1978 Earnings Variant I (α_t)	1978 Earnings Variant II $(\bar{C}\theta_t)$	1979 Earnings Variant I (α_t)	1979 Earnings Variant II $(\bar{C}\theta_t)$
Nonexperimental Estimates				
Linear Control Function Estimates:				
No Control Variables	157 (164)		79 (155)	
B1	686 (192)	726 (194)	494 (193)	534 (195)
B1+B2	231 (282)	638 (358)	-195 (286)	546 (360)
B1+W1	653 (203)	715 (260)	496 (230)	370 (289)
B1+B2+W1+W2	937 (263)	907 (335)	441 (303)	586 (386)
Weighted Average of Estimates	374 (146)		238 (152)	
Fixed Effect Estimates Constructed with t′=1972 Pre-Training Earnings:				
No Control Variables	231 (152)		153 (156)	
B1	699 (185)	736 (188)	508 (193)	544 (195)
B1+B2	938 (275)	1124 (353)	512 (287)	1032 (362)
Fixed Effect Estimates Constructed with t′=1974 Pre-Training Earnings:				
No Control Variables	475 (135)		397 (150)	
B1	713 (168)	693 (170)	522 (179)	500 (184)
B1+B2	946 (250)	800 (321)	520 (268)	708 (328)
Random Growth Estimates Constructed with t′=1973, t′-1=1972 Pre-Training Earnings:				
No Control Variables	494 (367)		460 (433)	
B1	78 (463)	44 (473)	-217 (546)	-263 (557)
B1+B2	486 (692)	-936 (898)	-15 (816)	-1372 (1965)
Random Growth Estimates Constructed with t′=1974, t′-1=1973 Pre-Training Earnings:				
No Control Variables	1276 (356)		1398 (471)	
B1	1183 (453)	981 (453)	1109 (576)	860 (576)
B1+B2	1278 (677)	880 (869)	935 (778)	808 (918)
Experimental Estimates				
	440 (142)		267 (162)	

NOTE: Standard Errors in Parentheses.

Table 5. Specification Tests of Nonexperimental Estimators for High School Dropouts (Youth)

Control Variable Set	Probability Values for: Pre-Program Tests Using Pre-Program Earnings		Overidentification Tests Using Pre-Program Earnings		Overidentification Tests Using 1978 Earnings		Overidentification Tests Using 1979 Earnings		Post-Program Tests Using 1978 Earnings		Post-Program Tests Using 1979 Earnings	
	$\alpha_t=0$	$\bar{C}_i\theta_t=0$	$\alpha_t=0$	$\bar{C}_i\theta_t=0$	$\alpha_t=0$	$\bar{C}_i\theta_t=0$	$\alpha_t=0$	$\bar{C}_i\theta_t=0$	$\alpha_t=0$	$\bar{C}_i\theta_t=0$	$\alpha_t=0$	$\bar{C}_i\theta_t=0$
Linear Control Function Estimators Constructed Using:												
	1975 Earnings as Dependent Variable				1978 or 1979 Earnings as Dependent Variable							
No Control Variables	0.000								0.000		0.000	
B1	0.000	0.000							0.000	0.000	0.000	0.000
B1+B2	0.000	0.012							0.000	0.000	0.000	0.000
B1+W1	0.000	0.208							0.000	0.012	0.000	0.000
B1+B2+W1+W2	0.016	0.336							0.005	0.632	0.033	0.000
Fixed Effect Estimators Constructed Using:												
	t=1974 and t'=1972 Earnings				t=1978 or 1979 and t'=1972 Earnings							
No Control Variables	0.000		0.000		0.000		0.000		0.000		0.000	
B1	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
B1+B2	0.000	0.019	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Fixed Effect Estimators Constructed Using:												
	t=1975 and t'=1972 Earnings				t=1978 or 1979 and t'=1974 Earnings							
No Control Variables	0.000		0.000		0.502		0.715		0.000		0.000	
B1	0.000	0.000	0.000	0.000	0.076	0.081	0.661	0.629	0.000	0.002	0.000	0.001
B1+B2	0.000	0.000	0.000	0.000	0.023	0.021	0.524	0.817	0.000	0.002	0.000	0.001
Random Growth Estimators Constructed Using:												
	t=1975, t'=1973, t'-1=1972 Earnings				t=1978 or 1979, t'=1973, t'-1=1972 Earnings							
No Control Variables	0.000		0.035		0.000		0.000		0.007		0.042	
B1	0.375	0.173	0.316	0.080	0.000	0.000	0.000	0.000	0.329	0.113	0.427	0.443
B1+B2	0.558	0.128	0.614	0.042	0.000	0.000	0.000	0.000	0.798	0.695	0.622	0.608
Random Growth Estimators Constructed Using:												
	t=1975, t'=1974, t'-1=1973 Earnings				t=1978 or 1979, t'=1974, t'-1=1973 Earnings							
No Control Variables	0.000		0.126		0.000		0.003		0.139		0.474	
B1	0.301	0.172	0.809	0.817	0.090	0.105	0.281	0.353	0.567	0.398	0.821	0.417
B1+B2	0.352	0.121	0.909	0.659	0.070	0.146	0.276	0.546	0.696	0.312	0.698	0.169

Table 6. Specification Tests of Modified Random Growth Estimator for High School Dropouts (Youth)

Control Variable Set	Overidentification Tests Using 1978 Earnings $\alpha_t=0$	$\bar{C}_i\theta_t=0$	Overidentification Tests Using 1979 Earnings $\alpha_t=0$	$\bar{C}_i\theta_t=0$	Post-Program Tests Using 1978 Earnings $\alpha_t=0$	$\bar{C}_i\theta_t=0$	Post-Program Tests Using 1979 Earnings $\alpha_t=0$	$\bar{C}_i\theta_t=0$
Modified Random Growth Estimators Constructed Using t'=1973, t'-1=1972 Pre-Training Earnings:								
B1	0.065	0.050	0.063	0.040	0.333	0.401	0.582	0.658
Modified Random Growth Estimators Constructed Using t'=1974, t'-1=1973 Pre-Training Earnings:								
B1	0.095	0.137	0.321	0.447	0.881	0.696	0.948	0.840

(Probability Values for: all columns)

Table 7. Estimates of Training Effects Using Modified Random Growth Estimators for High School Dropouts (Youths)

Control Variable Set	1978 Earnings Variant I α_t	$\omega_{t',t'-1}$	1978 Earnings Variant II $\bar{C}_i\theta_t$	$\omega_{t',t'-1}$	1979 Earnings Variant I α_t	$\omega_{t',t'-1}$	1979 Earnings Variant II $\bar{C}_i\theta_t$	$\omega_{t',t'-1}$
Modified Random Growth Estimators Constructed with t'=1973, t'-1=1974 Pre-Training Earnings:								
B1	-191	5.836	-183	5.942	-277	6.749	-270	6.927
	(329)	(1.021)	(298)	(1.253)	(351)	(.918)	(379)	(1.124)
Modified Random Growth Estimators Constructed Using with t'=1974, t'-1=1973 Pre-Training Earnings								
B1	-237	4.691	-201	4.573	-237	5.213	-213	5.011
	(367)	(1.001)	(361)	(1.116)	(385)	(1.023)	(370)	(.927)

NOTE: Standard Errors in Parentheses.

Table 8. Specification Tests of Nonexperimental Estimators for AFDC Recipients (Women)

Control Variable Set	Pre-Program Tests Using Pre-Program Earnings		Probability Values for: Overidentification Tests Using Pre-Program Earnings		1978 Earnings		1979 Earnings		Post-Program Tests Using 1978 Earnings		1979 Earnings	
	$\alpha_t=0$	$\bar{C}_i\theta_t=0$	$\alpha_t=0$	$\bar{C}_i\theta_t=0$	$\alpha_t=0$	$\bar{C}_i\theta_t=0$	$\alpha_t=0$	$\bar{C}_i\theta_t=0$	$\alpha_t=0$	$\bar{C}_i\theta_t=0$	$\alpha_t=0$	$\bar{C}_i\theta_t=0$
Linear Control Function Estimators Constructed Using:												
	1975 Earnings as Dependent Variable				1978 or 1979 Earnings as Dependent Variable							
No Control Variables	0.000								0.082		0.246	
B1	0.274	0.817							0.196	0.098	0.210	0.121
B1+B2	0.000	0.436							0.217	0.404	0.192	0.265
B1+W1	0.199	0.021							0.204	0.358	0.402	0.806
B1+B2+W1+W2	0.139	0.010							0.435	0.232	0.973	0.429
Fixed Effect Estimators Constructed Using:												
	t=1974 and t'=1972 Earnings				t=1978 or 1979 and t'=1972 Earnings							
No Control Variables	0.000		0.000		0.000		0.000		0.031		0.141	
B1	0.608	0.622	0.000	0.000	0.000	0.000	0.000	0.000	0.455	0.251	0.490	0.311
B1+B2	0.561	0.131	0.000	0.000	0.000	0.000	0.000	0.000	0.837	0.074	0.898	0.045
Fixed Effect Estimators Constructed Using:												
	t=1975 and t'=1972 Earnings				t=1978 or 1979 and t'=1974 Earnings							
No Control Variables	0.000		0.000		0.144		0.014		0.574		0.893	
B1	0.128	0.824	0.000	0.000	0.022	0.047	0.004	0.009	0.383	0.340	0.430	0.423
B1+B2	0.307	0.299	0.000	0.000	0.019	0.014	0.001	0.001	0.701	0.402	0.772	0.289
Random Growth Estimators Constructed Using:												
	t=1975, t'=1973, t'-1=1972 Earnings				t=1978 or 1979, t'=1973, t'-1=1972 Earnings							
No Control Variables	0.021		0.000		0.000		0.000		0.738		0.989	
B1	0.016	0.055	0.000	0.000	0.000	0.000	0.000	0.000	0.227	0.243	0.208	0.208
B1+B2	0.183	0.069	0.000	0.000	0.000	0.000	0.000	0.000	0.375	0.074	0.358	0.075
Random Growth Estimators Constructed Using:												
	t=1975, t'=1974, t'-1=1973 Earnings				t=1978 or 1979, t'=1974, t'-1=1973 Earnings							
No Control Variables	0.999		0.102		0.002		0.004		0.022		0.008	
B1	0.827	0.974	0.033	0.021	0.000	0.000	0.000	0.000	0.124	0.267	0.135	0.303
B1+B2	0.686	0.985	0.040	0.037	0.000	0.000	0.000	0.000	0.267	0.659	0.277	0.627