TECHNICAL WORKING PAPER SERIES

ACCOUNTING FOR DROPOUTS IN EVALUATIONS OF SOCIAL EXPERIMENTS

James Heckman
Jeffrey Smith
Christopher Taber

Technical Working Paper No. 166

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
September 1994

We have benefitted from comments at a demography workshop at the University of Chicago in January, 1994 sponsored by the Population Research Center, chaired by V.J. Hotz. This research was supported by NSF SBR-91-11455, NSF SBR-93-9183-45-36, by a grant from the Russell Sage Foundation, and by a grant from the Lynde and Harry Bradley Foundation of Milwaukee, Wisconsin. This paper is part of NBER's research program in Labor Studies. Any opinions expressed are those of the authors and not those of the National Bureau of Economic Research.

NBER Technical Working Paper #166
September 1994

ACCOUNTING FOR DROPOUTS IN EVALUATIONS OF SOCIAL EXPERIMENTS

ABSTRACT

This paper considers the statistical and economic justification for one widely-used method of adjusting data from social experiments to account for dropping-out behavior due to Bloom (1984). We generalize the method to apply to distributions not just means, and present tests of the key identifying assumption in this context. A reanalysis of the National JTPA experiment base vindicates application of Bloom's method in this context.

James Heckman
Department of Economics
University of Chicago
1126 East 59th Street
Chicago, IL 60637

Jeffrey Smith
Department of Economics
University of Western Ontario
London, Ontario N6A 5C2
CANADA

Christopher Taber
Center for Social Program Evaluation
University of Chicago
1126 East 59th Street
Chicago, IL 60637

I. Introduction

Bloom (1984) proposes a simple estimator of the mean impact of treatment on the treated in the context of an experimental evaluation in which some of those assigned to the experimental treatment group drop out prior to receiving treatment.[1] Many recent experimental evaluations (e.g., Bloom et al., 1993) of government social programs use this estimator to "adjust" for the effects of attrition from the program after random assignment. Despite its widespread use, little attention has been paid to several important aspects of this estimator.

In this paper, we (1) define the key identifying assumption underlying the estimator, (2) describe the estimator and its variance, (3) investigate the effect of failing to account for the estimation of the dropout rate in estimating the variance of the estimator, (4) generalize the estimator to the full distribution of outcomes for treatment group participants, (5) discuss conditions under which the identifying assumption is likely to be violated empirically, (6) present evidence about the failure of the assumption from a recent social experiment, (7) develop statistical tests of whether or not the identifying assumption holds, and (8) present alternative approaches to identification of treatment effects in the presence of dropping-out including two based on assumptions about earnings patterns within the treatment or control groups and others based on instrumental variables methods. We also (9) apply our procedure to analyze the 18 month outcomes produced from the National JTPA experiment that have been analyzed using Bloom's procedure. Our empirical analysis vindicates the application of Bloom's procedure in this instance. The paper ends with a summary and conclusion, wherein we emphasize the importance of letting the available evidence about the context of an evaluation guide the choice of an estimator.

II. The Identifying Assumption and the Corresponding Estimator

A. The Problem Posed by Dropouts and the Parameter of Interest

Consider an experimental evaluation in which persons who apply and are accepted into a program are randomly assigned into a treatment group eligible to receive the treatment and a control group ineligible to receive it. Assume further that there is no randomization bias, so that randomization does

---

[1] In the economics literature, this estimator was also proposed by Dubin and Rivers (1993) and Angrist and Imbens (1991). In the clinical trials literature, Haynes and Dantes (1987) discuss the problem of non-receipt of treatment by experimental treatment group members but do not propose a formal solution.

not alter behavior, that the treatment (or close substitutes for it) are not available from other sources, and that everyone in the treatment group receives the treatment.[2] The mean impact of the treatment on some outcome Y is defined as

$$\Delta = E(Y_t) - E(Y_c) \tag{1}$$

where Δ is the mean impact, E denotes mathematical expectation, $E(Y_t)$ is the mean outcome in the treatment group, and $E(Y_c)$ is the mean outcome in the control group. We leave implicit the conditioning on the event that all persons in the sample have applied to and have been accepted into the program as well as other conditioning variables. An unbiased estimator of Δ is

$$\bar{\Delta} = \bar{Y}_t - \bar{Y}_c \tag{2}$$

where "-" denotes sample mean. Estimates based on (2) can be constructed using post-treatment outcome data on those randomly assigned.

Now suppose that some persons in the experimental treatment group actually do not receive the treatment. Define d to be an indicator of this status, so $d = 1$ for a treatment group member who drops out of the program and $d = 0$ for a treatment group member who receives the treatment. Let $E(Y_t \mid d = 1)$ denote the mean outcome of the treatment group members not receiving the treatment (the dropouts), $E(Y_t \mid d = 0)$ denote the mean outcome of those receiving the treatment (the participants), and k_t denote the fraction of dropouts in the treatment group. Then the mean outcome in the treatment group may be decomposed into two components:

$$E(Y_t) = k_t\, E(Y_t \mid d = 1) + (1-k_t)\, E(Y_t \mid d = 0). \tag{3}$$

Random variables d are in principle also defined for control group members so that $d = 1$ if a control group member would have dropped out had he/she been a member of the treatment group and $d = 0$ is the complementary event. The control group mean outcome can be decomposed into

$$E(Y_c) = k_c\, E(Y_c|d = 1) + (1 - k_c)E(Y_c|d = 0) \tag{4}$$

where $E(Y_c \mid d = 1)$ denotes the mean outcome of those control group members who would have been dropouts, had they been in the treatment group, $E(Y_c \mid d = 0)$ denotes the mean outcome of the control group members who would have been participants, had they been in the treatment group, and k_c denotes

[2] See Heckman (1992) and Heckman and Smith (1993) for a discussion of these issues.

the fraction of controls who would have been dropouts. Obviously, as it stands decomposition (4) is not operational because we do not know d for control group members.

When there are dropouts from the experiment, the estimator defined in (2) provides an estimate of the mean impact of the availability of the treatment[3] rather than an estimate of the mean impact of the treatment on the treated, where this latter quantity is defined as

$$\Delta_p = E(Y_t \mid d = 0) - E(Y_c \mid d = 0). \tag{5}$$

The parameter Δ_p is sometimes viewed to be of greater interest than the mean impact of treatment availability. It informs us about the difference in outcomes between persons who obtain "full" treatment and those who receive none at all. Interest in this parameter is based on the assumptions that (a) dropouts receive no "treatment" at all or (b) that the partial treatment received by dropouts is not of primary interest and does not inform analysts about the impact of full treatment or (c) that we may want to distinguish partial treatment effects from full treatment effects. In the last instance we might also be interested not only in Δ_p but also in the analogous parameter for the individuals who receive partial treatment. More generally, we may have two treatment streams and we would like to distinguish separate treatment effects for each stream as in Hotz and Sanders 1994. We discuss these issues below. We note in passing, however, that Δ is often the parameter of interest in performing marginal evaluations of existing programs taken as a whole inclusive of the dropping-out behavior of participants.

Data on the post-treatment outcomes of those in the treatment and control groups are not enough to estimate (5) because $E(Y_c \mid d = 0)$ is not identified. Outcome data alone provides no way to sort the control group into those who would and would not have been dropouts. Additional assumptions, or additional data, are required.

B. Bloom's Identifying Assumption

Bloom (1984) assumes that the mean outcome of the dropouts in the treatment group is the same as the mean outcome of the persons in the control group who would have been dropouts, had they been in the treatment group. More formally, he assumes that

---

[3]This parameter is often referred to as "intent to treat" in the biostatistics literature. See, e.g., Efron and Feldman (1991).

(A-1) $$E(Y_t \mid d = 1) = E(Y_c \mid d = 1).$$

Only equality of the means is required. Variances and other parameters of the conditional distributions $(Y_t \mid d = 1)$ and $(Y_c \mid d = 1)$ need not be the same.

To see how this assumption provides the necessary identification, first note that random assignment implies that

(6) $$k_t = k_c = k.$$

Combining this fact with assumption (A-1) allows (4) to be solved for $E(Y_c \mid d = 0)$

(7) $$E(Y_c \mid d = 0) = [1/(1-k)]\, E(Y_c) - [k/(1-k)]\, E(Y_t \mid d = 1).$$

Substituting (7) into (5) yields,

(8) $$\Delta_p = E(Y_t \mid d = 0) - [1/(1-k)]\, E(Y_c) + [k/(1-k)]\, E(Y_t \mid d = 1).$$

Solving (3) for $E(Y_t \mid d = 0)$ and substituting into (8) we obtain

(9) $$\Delta_p = [1/1-k]\, [E(Y_t) - E(Y_c)] = [1/1-k]\, \Delta.$$

The corresponding estimator is

(10) $$\bar{\Delta}_p = [1/(1-k)]\, (\bar{Y}_t - \bar{Y}_c) = [1/(1-k)]\, \bar{\Delta}\,.$$

Writing the estimator in this form shows that it scales up the estimate of the mean impact of treatment availability into an estimate of the mean impact of participation by assuming a zero impact for the dropouts. Since in practice k is estimated rather than known, the feasible form of the estimator replaces k with sample proportion $\hat{k}_t$ obtained from the treatment group.

D. Variance of the Mean Impact Estimator

In the analysis of the JTPA experiment, (see, e.g., Bloom et al., 1993) the authors calculate the standard errors for impact estimates obtained using the estimator in (10) under the assumption that the dropout rate k is known with certainty. Under this assumption, the variance of the estimator in (10) is:

(11) $$\text{Var}(\bar{\Delta}_p) = \text{Var}(\bar{Y}_t | d = 0) + (1 / (1 - k))^2 [\text{Var}(\bar{Y}_c) + k^2 \text{Var}(\bar{Y}_t | d = 1)]$$
$$= [\text{Var}(\bar{Y}_t) + \text{Var}(\bar{Y}_c)] / [1-k]^2$$

In this section, we investigate the extent of the downward bias in the estimated covariance of the impact estimates induced by the assumption that k is known. A first order asymptotic approximation to the estimator defined in (10) yields the following additional term C that serves to correct the variance given in (11) for the estimation of k:

(12) $$C = [[E(Y_t) - E(Y_c)] / (1 - k)^2]^2 \frac{k(1 - k)}{N_t}$$

where N_t is the number of people in the treatment group. Both the uncorrected variance in (11) and the correction term C can be estimated using the sample analogs of $E(Y_c)$, $E(Y_t)$ and k. Observe that the closer the control group mean to the treatment mean, the smaller the proportion of dropouts, and the larger N_t, the smaller the correction C.

Table 1 displays the ratio of the estimated variances of the estimated impact of treatment on the treated in the cases where k is not and is assumed to be known with certainty for various values of k and various assumed mean differences in outcomes between the control and dropout groups. This ratio is calculated using earnings in the 18 months after random assignment for adult males in the recent experimental evaluation of the employment and training programs provided under the Job Training Partnership Act (JTPA). This evaluation is denoted the National JTPA Study (NJS) and is described in more detail below.

As is plain from (12), the ratio of corrected to uncorrected variances is 1.00 when the difference in means is zero, because in this case the correction term is also zero. The ratio increases as the mean difference in outcomes between the two groups increases. It also increases with k from 0.0 to 0.5 and then decreases symmetrically as k goes from 0.5 to 1. In none of the cases shown does the correction ever increase the estimated variance by more than one percent.

Table 2 translates the importance of the correction into dollar terms by displaying the difference between the estimated standard error of the impact of treatment on the treated with and without the correction for the estimation of the dropout rate. The same basic patterns found for the ratio hold for this quantity as well. The main difference is that the sample size, which did not affect the variance ratio, does affect the difference in estimated standard errors. As expected, the correction matters less the larger the sample. The greatest effect of the correction shown in Table 2 is $11.49 in the case of a sample of

300 treatment group members, a dropout rate of 50 percent and a mean earnings difference of $2000 between the control and treatment groups. While not trivial, even this difference is unlikely to constitute more than a small fraction of the uncorrected standard error of the impact estimate.

The basic conclusion from this section is that for the JTPA experiment correcting for the estimation of the dropout rate is unlikely to have a meaningful effect on the size of the estimated standard errors of estimates of the impact of treatment on the treated obtained using the estimator (10). As a result, it is also unlikely to affect statistical conclusions reached under the assumption that the dropout rate is known with certainty. For this reason, we assume that k is known throughout the remainder of the paper in our analysis of JTPA data.

E. Generalization to Full Distributions

While Bloom (1984) and Dubin and Rivers (1993) confine their attention to mean impacts, both the identifying assumption in (A-1) and the resulting estimator generalize to allow the full distribution of outcomes of treatment group participants to be obtained. Let $F_t(y_t \mid d = 1)$ be the distribution of $(Y_t \mid d = 1)$ and let $F_c(y_c \mid d = 1)$ be the distribution of $(Y_c \mid d = 1)$, then the distributional analog to (A-1) is:

$$F_t(y \mid d = 1) = F_c(y \mid d = 1). \tag{A-2}$$

Under this assumption, the outcome distribution for members of the control group who would have received treatment had they been in the treatment group is given by:

$$F_c(y \mid d = 0) = [1/(1-k)]\, F_c(y) - [k/(1-k)]\, F_t(y \mid d = 1), \tag{13}$$

where $F_t(y \mid d = 1)$ and $F_c(y)$ are the cumulative distribution functions of $(Y_t \mid d = 1)$ and Y_c respectively.

Any identifying assumption is based on an implicit or explicit model of behavior for the dropouts. Clarity and scientific rigor are served by making explicit the implicit assumption that justifies (A-1). The assumption rules out any real effect of participating in the experiment on dropouts. Its plausibility can only be judged in the context in which it is invoked. This paper considers the case of job training programs, where the estimator in (10) has been extensively applied.

For two reasons, assumption (A-2) is useful. First, if we are interested in any attribute of the distribution of Y_c given $d = 0$, it will be identified from equation (13). Furthermore, by replacing $F_c(y \mid d = 0)$ and $F_t(y \mid d = 1)$ by their sample counterparts in equation (13), we can obtain a consistent

estimate of $F_c(y \mid d = 0)$. The sample counterparts of these distributions are empirical distribution functions which converge asymptotically to Gaussian processes, so it is easy to show that the estimated $F_t(y \mid d = 1)$ converges asymptotically to a Gaussian process.

The second reason (A-2) is useful is that it provides a testable restriction which we will exploit below.

III. Cases Where the Identifying Assumption Will Not Hold

A. The Role of Partial Treatment

For certain types of programs, assumptions (A-1) and (A-2) are plausible. In drug trials, if dropouts leave the program before receiving any dose of the drug, it is unlikely that any treatment effect exists for them. However, if some or all dropouts receive partial treatment, their mean treatment effect is likely to be non-zero, unless there is a threshold level below which the drug has no effect and dropouts are invariably below the threshold.

It is clarifying to extend Bloom's framework and introduce three latent random outcome variables for each person. Y_p is the outcome the person would receive if he/she receives full treatment; Y_d is the outcome a person would receive if he/she receives partial treatment; Y_c is the outcome the person would receive if he/she does not participate at all. "Partial treatment" may designate a state in which a person is randomized into the program but receives no treatment.

For clarity, we postulate an underlying model of outcomes (Y_d, Y_p, Y_c, d) in which each person's outcome is a function of their type θ, so that $(Y_d, Y_p, Y_c, d) = (Y_d(\theta), Y_p(\theta)\ Y_c(\theta), d(\theta))$.[4] We assume that there is a distribution of θ types in the population. In the sub-population of persons who are eligible for the treatment and who wish to participate, not all θ types may be represented. The distribution of θ in this sub-population is $F(\theta)$, where θ may be a vector or a scalar.

Treatment alters the base (control) state outcome by an amount $\alpha(\theta)$. Partial treatment alters $\alpha(\theta)$. We denote partial treatment effects by $\nu(\theta)\alpha(\theta)$, where $\nu(\theta)$ is an adjustment factor. We may think of $\nu(\theta)$ ϵ [0,1], but this is not necessary. Thus we have a model in which, for the population of persons who are drawn from $F(\theta)$,

---

[4]Formally we are assuming that (Y_d, Y_p, Y_c, d) is measurable with respect to the sigma algebra generated by θ.

(14a) $$Y_c = Y(\theta)$$

(14b) $$Y_d = Y(\theta) + \alpha(\theta)\nu(\theta)$$

(14c) $$Y_p = Y(\theta) + \alpha(\theta).$$

We assume that individual differences in θ lead some persons to drop out and others to stay in the program.

Let Θ_d denote the set of θ values of persons who drop out. Formally,

$$\Theta_d = \{\theta \mid \text{persons of type } \theta \text{ drop out}\}.$$

Note that $\Pr(\theta \in \Theta_d) = k$ is the probability that a person drops out. Then

$$E(Y_c) = E(Y(\theta))$$

and

$$E(Y_d) = E[Y_d \mid \theta \in \Theta_d]\Pr(\theta \in \Theta_d) + E[Y_p \mid \theta \notin \Theta_d]\Pr(\theta \notin \Theta_d).$$

Thus $$\Delta_p = E(\alpha(\theta) \mid \theta \notin \Theta_d).$$

The impact of full treatment on the fully treated is just the expected value of the treatment effect $\alpha(\theta)$ for persons whose θ's lead them <u>not</u> to drop out. Substituting into (10) and simplifying, we obtain:

(15) $$E(\bar{\Delta}_p) = \Delta_p + [k/1\text{-}k]\, E[\alpha(\theta)\nu(\theta) \mid \theta \in \Theta_d].$$

If there is no partial effect of treatment for the dropouts, $\nu(\theta) \equiv 0$ for $\theta \in \Theta_d$ and $\bar{\Delta}_p$ is unbiased for Δ_p. This assumption of no partial treatment for dropouts is the key condition that justifies Bloom's estimator.

From (14b) we see that $\bar{\Delta}_p$ may be unbiased for Δ_p if $\alpha(\theta)$ and $\nu(\theta)$ possess just the right covariance. That is, if

$$\mathrm{Cov}(\alpha(\theta), \nu(\theta) \mid \theta \in \Theta_d) = -\, E[\alpha(\theta) \mid \theta \in \Theta_d]\, E[\nu(\theta) \mid \theta \in \Theta_d],$$

then the expectation of the second term on the right hand side of (14b) must be zero. If $E(\alpha(\theta) \mid \theta \in \Theta_d) \geq 0$, then the length of time spent in the program and the program impact have to be negatively correlated <u>among</u> dropouts in order for the final term in (15) to be zero. However, if program participation is based on expected gain, it is plausible that $E(\alpha(\theta) \mid \theta \in \Theta_d) < 0$ since persons with $\theta \in \Theta_d$ choose not to attend. In this case, a sufficient condition for no bias is that α and ν are positively

correlated for $\theta \in \Theta_d$.

In this setup, $\alpha(\theta)$ may be a pure endowment or a stigma effect. Persons who get into a program may carry a label that helps or harms them. Burtless (1985) presents convincing evidence of tangible harm that results from being associated with a job subsidy program. Since $\alpha(\theta)$ is defined relative to $Y(\theta)$, the label may also be attached to those randomized out instead of those randomized in. The distinction is purely semantic.

If the positive or negative endowment is the same for all persons who are randomized into the program, irrespective of their time in the program, then $\nu(\theta) \equiv 1$ and

$$E(\bar{\Delta}_p) = E(\alpha(\theta))\ [1/(1\text{-}k)] \neq \Delta_p.$$

If the dropouts are the high α persons, then the estimator is upward biased and perhaps badly so. If the dropouts are the low α persons, the bias is ambiguous as the underestimate of the expected value term acts in the opposite direction as the re-scaling by [1/(1-k)]. This example again raises the interesting question of why Δ_p should be the parameter of interest. We discuss this issue below. First we consider evidence from an experimental evaluation of an employment and training program on the question of whether or not $\nu(\theta) \equiv 0$.

B. Evidence on Partial Treatment from the National JTPA Study

In this section, we present empirical evidence on the importance of bias in the Bloom estimator in the context of the National JTPA Study (NJS). This is the recent experimental evaluation of JTPA already introduced above. The JTPA program provides basic education, classroom training in occupational skills, subsidized on-the-job training at private firms, job search assistance, and other employment and training services to economically disadvantaged persons. The program includes a performance standards system in which locally managed training sites compete for incentive payments based on their success at placing enrollees in steady, high-paying jobs.

In the NJS, random assignment occurred prior to formal enrollment in the program. A substantial fraction of those randomly assigned to the experimental treatment group never enrolled in the program. As shown in Table 3, over 37 percent of adult male treatment group members did not enroll in JTPA during the 18 months following random assignment. Similar non-enrollment rates are reported for the other three target groups in the NJS. These non-enrollment percentages are used to generate estimates of Δ_p using the estimator in (10) by Bloom et al. (1993).

This high rate of non-enrollment results in part from the time lag that often occurs between

random assignment and the initiation of training. For courses given on an academic schedule, the applicant must wait until the beginning of the next quarter or semester. During this waiting period, the applicant may find a job or lose interest, and so may fail to enroll. However, as long as the training and job seeking activities of randomized-in potential trainees are the same as those of randomized-out potential trainees of the same type, Bloom's estimator of Δ_p remains valid. If, however, access to the program affects the outcome measure, it is not clear that Δ_p is the parameter of interest.

A more serious problem results from the mechanics of the JTPA performance standards system. Incentive payments to the training sites under the JTPA performance standards system depend only on the performance of their enrollees. At the same time, enrollment in JTPA is very flexible, so that sites can often delay enrolling someone until it is clear that the person is likely to succeed in training. For trainees assigned to job search assistance or to on-the-job training, this can mean that enrollment occurs when they find a job or an employer willing to provide them with on-the-job training. Those not enrolled for this reason are counted as dropouts, even though they often receive assistance in looking for jobs, writing resumes and presenting themselves in interviews and may acquire information about the local labor market not available to persons randomized out. These activities would increase their future earnings even if they do not find a job or an on-the-job training slot during the period of their contact with the JTPA program.

Table 4 shows the extent of JTPA contact following random assignment among a subset of the treatment group non-enrollees. Over half of this subset of non-enrollees received some JTPA services. This evidence suggests that assumption (A-1) may be inappropriate, as those in the control group who would have been dropouts had they been in the treatment group did not receive these services. Many of these JTPA dropouts are not "no shows" who receive no treatment from the program. Below, we present empirical tests of (A-2) using data from the JTPA evaluation. In the next section, we explore the possible consequences of violations of (A-1) for the impact estimates obtained in the JTPA evaluation.

C. Sensitivity of Impact Estimates to Violations of (A-1)

To explore the sensitivity of the estimated experimental impacts to departures from (A-1), define ε as $\varepsilon = E(Y_t \mid d = 1) - E(Y_c \mid d = 1)$. Then a simple modification of the derivation leading up to (9) produces an adjusted version of Δ_p

(16) $$\bar{\Delta}_p^{\varepsilon} = \bar{\Delta}_p - [k/(1-k)]\varepsilon \ .$$

Table 5 displays $\bar{\Delta}_p^{\varepsilon}$ calculated with ε equal to \$150, \$100, \$50, -\$50, -\$100, and -\$150. As in Table 1, the outcome variable Y corresponds to the sum of self-reported earnings in the 18 months after random assignment. The estimated means and sample sizes underlying these values appear in Table 3. As noted above, in the context of the JTPA evaluation, the most likely source of bias is receipt of partial treatment by some of the treatment group dropouts. Suppose that the mean impact of these partial services on earnings in the 18 months after random assignment is \$300, which is quite reasonable given the annual impact findings for job search assistance reported in Gueron and Pauly (1991). As the evidence in Table 4 indicates that roughly half of the dropouts received some partial treatment, this corresponds to a value of ε equal to -\$150. For adult males, ε = -\$150 implies an adjusted earnings impact estimate of $\bar{\Delta}_p^{\varepsilon}$ = \$812.77. The figure in square brackets indicates that this estimate is \$88.78, or around 10 percent, lower than the estimate given by the Bloom estimator.

D. The Parameter of Interest in the Presence of Partial Treatment

Our discussion of stigma and endowment effects suggests that simply being randomized into a treatment group may have real effects on outcomes. In the special case of $\nu(\theta) \equiv 1$, the correct definition of the treatment effect is not Δ_p but Δ. More generally, one might wish to distinguish the effects of partial exposure ($0 < \nu(\theta) < 1$) from those resulting from full exposure ($\nu(\theta) = 1$) even though both are consequences of the program. In these cases, it is inappropriate to ignore the benefits of partial participation as is done when Δ_p is defined to be the parameter of interest.

A stark example of this point is given by the following example. Consider a job subsidy program, in which each person in the treatment group receives a base gain $\alpha(\theta)$, and where alternative market offers arrive after the time of randomization. If the market offer $Y(\theta) + \eta(\theta)$ exceeds $Y(\theta) + \alpha(\theta)$, then the person accepts it and becomes a "no show" or dropout. Dropouts will generally earn more on average than program participants. In this context, it is not obvious that Δ_p is an economically meaningful measure of program impact.

Persons randomized into treatment have expected outcomes of

$$E[\text{Max}(Y(\theta) + \alpha(\theta); Y(\theta) + \eta(\theta))]$$

whereas persons who are randomized out have one fewer option and have expected outcome

$$E[Y(\theta) + \eta(\theta))].$$

The gain to the program comes from giving participants a base wage on which they can only improve.

In this situation Bloom's estimator for Δ_p is valid since

$$\begin{aligned}(17)\ E(\bar{\Delta}_p) = [1/(1\text{-}k)]\ [&k\ E(Y(\theta) + \eta(\theta) \mid \eta(\theta) \geq \alpha(\theta)) + (1\text{-}k)\ E(Y(\theta) + \alpha(\theta) \mid \alpha(\theta) > \eta(\theta)) \\ &- k\ E(Y(\theta) + \eta(\theta) \mid \eta(\theta) \geq \alpha(\theta)) - (1 - k)\ E(Y(\theta) + \eta(\theta) \mid \alpha(\theta) > \eta(\theta))] \\ &= E(\alpha(\theta) - \eta(\theta) \mid \alpha(\theta) \geq \eta(\theta)) \\ &= \Delta_p.\end{aligned}$$

Persons who drop out to get a higher private sector wage have counterparts in the control population who, for the same θ, have the same wage. Thus $E(Y_t \mid d = 1) = E(Y_c \mid d = 1)$ which is Bloom's assumption (A-1).

The relevant issue is whether Δ_p is an interesting parameter for evaluating the program in this context. It exaggerates the impact of the program because it measures only that part of the treatment attributable to those who do not receive higher private sector wage offers, ignoring the important point that k percent of the persons offered the subsidy would have received higher wages through the usual wage offer process even in the absence of the program.

If real resources are spent on all persons in the treatment group, say by subsidizing search, but only proportion (1-k) benefits from it, the appropriate per person gross marginal benefit for the program is

$$\Delta = (1 - k)\Delta_p.$$

On the other hand, if $\alpha(\theta)$ is really a wage subsidy given only to those who actually take it, Δ_p is the correct measure of per person gross benefits. The context of the program determines the appropriate parameter of interest and whether or not adjusting for dropouts is an economically interesting exercise even when it produces Δ_p.

IV. Testing Bloom's Assumption

A. A Testing Strategy

In this section, we outline a strategy for testing restrictions on the constructed outcome distribution of the participant analogs in the control group implied by assumption (A-2). While the estimator in (10) requires only the weaker assumption (A-1), this condition is implied by (A-2) and is unlikely to hold in the absence of (A-2) except by coincidence. Thus, we test restrictions on the distribution of outcomes implied by (A-2) for persons in the control group who would have participated (received treatment) had they been in the treatment group.

The restriction that we will test can be seen from equation (13) which is an immediate consequence of assumption (A-2). We know $F_c(y_c \mid d = 0)$ must be a legitimate distribution function, thus we obtain the following restriction on it:

$$(R - 1) \qquad F_c(y \mid d = 0) = \frac{F_c(y) - k\, F_t(y \mid d = 1)}{1 - k} \quad \textit{is a proper c.d.f.}$$

Any estimated distribution constructed from the empirical counterparts to the ingredients on the right of R-1 must satisfy this condition. This restriction imposes a number of testable restrictions on the generating distribution, such as

$$(R - 2) \qquad 0 \leq \frac{F_c(y) - kF_t(y \mid d=1)}{1 - k} \leq 1$$

for all Y and that the constructed cdf should be monotonically increasing in Y. In this section we will discuss tests of some of these restrictions.

A weakness of our proposed strategy is that any test of (A-2) based on (R-1) may not be consistent. (A test is consistent if as sample size becomes large, the power of the test goes to one). (A-2) is sufficient for (R-1) but not necessary. It is easy to construct examples of distribution functions that satisfy (R-1) but not (A-2). Any test based on (R-1) will have no power against such alternatives. Thus rejection of (R-1) constitutes rejection of (A-2), but acceptance of (R-1) is not evidence in support of (A-2).

The testing strategy outlined here begins with simple tests of differences between the dropout and control distributions. Under assumption (A-2), the control and dropout distributions will be equivalent if the factors causing persons to drop out are unrelated to their outcomes. If $F_c(y) = F_t(y \mid d = 1)$ then

restriction (R-1) will hold. In practice, when these two outcome distributions are not statistically distinguishable, the implied outcome distribution for the participant analogs in the control group is extremely unlikely to violate any of the restrictions tested by the more complicated tests proposed below. Thus, in cases where the initial battery of tests fails to reject the equivalence of the control and treatment group dropout outcome distributions, the additional tests proposed here should be superfluous. In results not reported here, we do not reject the hypothesis of equality of the two distributions for any demographic group. (We use both Kolmogorov-Smirnov and Wilcoxon tests). This provides some support for Bloom's assumption applied to the National JTPA data. Taken at face value, it suggests that dropping out from the program is random with respect to control state outcomes.

B. Tests Restrictions Implied by the Identifying Assumption

A simple test in contexts such as the JTPA experiment in which both the control and dropout distributions have substantial point masses at zero earnings (or some other specific outcome value), is to test the null hypothesis that the probability of zero earnings for the participant analogs calculated under (A-2) lies in [0,1]. We performed this test using the JTPA outcome data and failed to reject the null hypothesis.

More generally, the restriction (R-1) can be written in a different, but equivalent form,

$$(R\text{-}1') \qquad \frac{Pr(Y_c \in b) - kPr(Y_d \in b)}{1 - k} \geq 0 \quad \textit{for all } b \subset B,$$

where B is the union of the supports of Y_c and Y_d. For instance restriction (R-2) is a special case of (R-1') if we consider only the sets b of the form $(-\infty, Y)$ and (Y, ∞). Unless the support B is finite, testing (R - 1') for every subset of B is impossible. The tests in this subsection are devoted to testing (R - 1') for a finite number of subsets of B. In the case where B is finite we will have completely exhausted the implications of assumption (R-1). In the case where B is infinite, we could in principal develop a better test. However, the test we develop here is easy to implement in practice and has known asymptotic properties.

The null hypothesis we consider is

$$(R\text{-}1'') \qquad H_0: \frac{Pr(Y_c \in b_j) - kPr(Y_d \in b_j)}{1 - k} \geq 0 \quad \textit{for } j = 1,\ldots,J\ .$$

For each j we can estimate $(1/(1\text{-}k)[\Pr(Y_c \in b_j) - k \Pr(Y_d \in b_j)]$ by using sample averages. Define $\hat{P}$ to be the vector of these estimates and P to be the true value. It is easy to show that
where each component of the matrix. Σ is of the form,

$$\sqrt{n}\ (\hat{P} - P) \rightarrow N(0,\Sigma)\ ,$$

$$\sigma_{ij} = \frac{Pr(Y_c \in b_j \cup b_j) - Pr(Y_c \in b_i)Pr(Y_c \in b_j) + k^2[Pr(Y_d \in b_i \cup b_j) - Pr(Y_d \in b_i)Pr(Y_d \in b_j)]}{(1 - k)^2}.$$

We continue to assume that k is known. Modifying this result to account for estimation of k is straightforward.

Testing H_0 requires that we define a test statistic $t(\hat{P})$ and derive the asymptotic distribution of this test statistic under the null hypothesis. The composite nature of this null hypothesis complicates the procedure. There is no similar test region in this case. For any critical region we calculate the size of the test based on the least favorable distribution (e.g. the distribution consistent with the null hypothesis for which the probability of rejection is the greatest).

One test statistic is based on the Wald test. Following Wolak (1991) we can define

$$t(\hat{P}) = \inf_{M \geq 0} [(\hat{P} - M)' \hat{\Sigma} (\hat{P} - M)].$$

The distribution of this test statistic depends on the particular null under which we evaluate it. Wolak (1991) derives this distribution and provides some information about where the least favorable distribution will occur. He shows that at least two constraints in the finite version (R-1'') will bind.

Another possible test statistic is

$$t(\hat{P}) = \inf_{i=1,\ldots,J} \frac{\hat{P}_j}{\hat{\sigma}_{jj}}\ ,$$

which is just the minimum value of the t-statistic for each individual cell. In the appendix we develop this test explicitly and provide some theoretical and Monte Carlo evidence about where the least favorable distribution may lie. We show that typically as many of the constraints as possible will bind at the least favorable distribution. In application to the JTPA data, we do not reject the null hypothesis. (R-1'') so that Bloom's assumption receives support from this test.

IV. Alternative Identifying Assumptions

A. Assumptions about the Effectiveness of Partial Treatment

The simplest alternative to the estimator (10) applies in cases where the mean impact of partial treatment on the dropouts is known or can be reliably estimated. This condition is expressed formally as "ε is known" where

$$\varepsilon = E(Y_t \mid d = 1) - E(Y_c \mid d = 1). \tag{A-3}$$

Using (A-3) in place of (A-1) in the derivation of (10) gives the estimator in (16) above. The Bloom estimator (10) is the special case of this class of estimators where $\varepsilon = 0$.

B. Assumptions about Relative Outcomes Within the Control Group

In this subsection we consider a class of estimators for Δ_p that do not require any assumptions about the mean earnings of dropouts within the treatment group. These estimators represent alternatives to the Bloom estimator in situations where (A-1) does not hold, and where information on ε is not available.

Consider the simplest member of this class first. Assume that treatment group members drop out of the program at random, where we define dropping out at random to mean that

$$E(Y_c \mid d = 1) = E(Y_c \mid d = 0). \tag{A-4}$$

That is, we assume that the mean income of the control group members who would have been dropouts in the treatment group equals the mean income of the control group members who would have received the treatment had they been in the treatment group.

Assumption (A-4) identifies $E(Y_c \mid d = 0)$. Substituting (A-4) into (4) and solving yields

$$E(Y_c \mid d = 0) = E(Y_c). \tag{18}$$

An unbiased estimator of Δ_p is then

$$\bar{\Delta}_p = \bar{Y}_t(d = 0) - \bar{Y}_c \tag{19}$$

or the difference between the sample mean outcomes of the participants and the controls. ($\bar{Y}_t(d = 0)$ is the mean of treatment group members who are not dropouts.) This estimator makes no use of the sample

mean outcome of the treatment group dropouts. For this reason, it is robust to failures of (A-1).

The estimator defined by (19) belongs to a general class of estimators that rely on assumptions about the relative magnitudes of $E(Y_c \mid d = 1)$ and $E(Y_c \mid d = 0)$ to identify $E(Y_c \mid d = 0)$. A more general form of the identifying assumption for this class of estimators is

(A-5) $$E(Y_c \mid d = 1) = \eta E(Y_c \mid d = 0)$$

where each value of η, the constant of proportionality, defines a separate member of the class. Substituting (A-5) into (4) and solving for $E(Y_c \mid d = 0)$ yields

(20) $$E(Y_c \mid d = 0) = E(Y_c) / [1 + k(\eta-1)].$$

The value on the left hand side decreases as η increases so that the overall mean $E(Y_c)$ remains the same. Under (A-5), the general form of the estimator in (19) becomes

(21) $$\Delta_p^{\eta} = \bar{Y}_t(d = 0) - \bar{Y}_c(d = 0) = \bar{Y}_t(d = 0) - \{\bar{Y}_c(d = 0) / [1 + k(\eta-1)]\}.$$

In cases with substantial uncertainty about the appropriate value of η, presenting estimates based on several different values provides information about the sensitivity of the estimates to the choice of η and allows the reader to weight the estimates in accord with his or her own prior. Hotz and Sanders (1994) present another way of conducting sensitivity analyses in a more structured setting.

Table 6 presents estimates of Δ^{η}_p using data on self-reported earnings in the 18 months after random assignment from the National JTPA Study for various choices of η. The first line of the table repeats the estimates Δ_p obtained using the Bloom estimator. The next five rows present estimates Δ^{η}_p for values of η equal to 0.50, 0.75, 1.00, 1.25, and 1.50. The final row displays the value of η that equates Δ^{η}_p and Δ_p for each target group.

The table shows that for η in [0.5,1.5], varying the selection process into the dropout group by varying the value of η strongly influences the resulting impact estimates Δ^{η}_p. Setting η to 0.50, which implies that those controls who would have been dropouts have mean earnings only half that of those who would have been participants, produces strongly negative impact estimates in all cases. In contrast, η equal to 1.5, which implies a lower mean outcome for those who would have received the treatment than for those who would have dropped out, produces very large impact estimates in all cases. The estimates for $\eta = 1.0$, which corresponds to the random drop out case considered above, come close to those from the Bloom estimator for three of the four target groups, with male youth the exception. The bottom row of Table 6 provides additional evidence on this point, as the value of η that equates the two estimates lies

close to 1.0 except for male youth.

C. Exploiting The Information In Exclusion Restrictions

For a number of econometric models of self-selection, it has been shown that the use of exclusion restrictions is helpful for identification of the parameters of interest. (See e.g. Heckman 1990, Angrist and Imbens 1991, and Imbens and Angrist 1994). In this section we explore whether this approach is useful when assumption (A-1) fails to hold.

It is useful to distinguish between two possible parameters of interest. Thus far we have been concerned with the parameter

$$\Delta_p = E(Y_p \mid d = 0) - E(Y_c \mid d = 0),$$

where as above, in the context of the control group d is an indicator of an individual who would drop out of the program if they were randomized into the treatment group. This is the expected effect of participating in the program among those who participate. This is sometimes called the effect of treatment on the treated or more precisely the effect of full treatment on the fully treated. We contrast this with the parameter

$$\widetilde{\Delta}_p \equiv E(Y_p) - E(Y_c).$$

This is the expected effect of the program in the counterfactual state that nobody drops out of the program. This is the interesting parameter if coverage in the program becomes universal. The two parameters answer two different policy counterfactuals.

We define an exclusion restriction as some random variable Z with support $\underline{\underline{Z}}$ that influences the decision to drop out, but does not influence the outcomes. A standard restriction writes

(A-6)
$$E(Y_p \mid Z) = E(Y_p)$$
$$E(Y_d \mid Z) = E(Y_d)$$
$$E(Y_c \mid Z) = E(Y_c)$$
$$\Pr(d = 1 \mid Z) \neq \Pr(d = 1).$$

First consider identification of $\widetilde{\Delta}_p$. Clearly $E(Y_c)$ is identified from the control group alone. Identification of $E(Y_p)$ is achieved with assumption (A-1) and the additional assumption that for some $z \in Z$, $\Pr(d = 0 \mid Z = z) = 1$. In this case $E(Y_p \mid d = 0, Z = z) = E(Y_p \mid Z = z) = E(Y_p)$. This case is a special case of the identification theorems in Heckman (1990) or Angrist and Imbens (1991).

To estimate Δ_p note that one can easily estimate $E(Y_c)$ from data on controls, and one needs to use some selection-correction procedure to estimate $E(Y_p)$. (See, e.g. Amemiya, 1985, for a discussion of sample selection correction procedures).

In contrast assumption A-6 is not helpful for identification of the parameter Δ_p. The parameter $E(Y_p \mid d = 0)$ is identified from participant, earnings but the parameter $E(Y_c \mid d = 0)$ is not identified except in the limit set where $Pr(d = 0 \mid Z = z) = 1$, in which conditional versions of $\widetilde{\Delta}_p$ and Δ_p coincide. Since we never observe d for any control, assumption (A-6) is not helpful. This point is clarified in the following example. We demonstrate that even explicit modeling and strong functional form assumptions does not deliver identification of the parameter Δ_p under assumption (A-6).

Example 1: We consider the following special case:

$$Y_c = \mu_c + \varepsilon_c$$
$$Y_p = \mu_p + \varepsilon_p$$
$$Y_d = \mu_d + \varepsilon_d$$
$$d = 1(Z\gamma + \nu \geq 0)$$

where $1(\cdot)$ is the indicator function which takes the value 1 if its argument is true, and takes the value 0 if it is false. We assume $(\varepsilon_c, \varepsilon_p, \varepsilon_d, \nu)$ is a normal random vector that is independent of the scalar random varible Z. In addition we normalize the variance of ν to one. In this case we can write

$$E[Y_c \mid d = 0, Z] = E[Y_c \mid Z, Z\gamma + \nu < 0] = \mu_c + \sigma_{c\nu}\,\lambda(-Z\gamma)$$

where $\sigma_{c\nu}$ is the covariance between ν and ε_c, and $\lambda(\cdot)$ is the inverse Mills ratio. Since we do not observe d for the control group, without further assumptions we cannot identify $\sigma_{c\nu}$ and thus we can not identify

$$\Delta_p(Z) = \mu_p - \mu_c + (\sigma_{p\nu} - \sigma_{c\nu})\lambda(-Z\gamma)$$

where $\sigma_{p\nu}$ is the covariance between ν and ε_p. Note, however, that we can identify $\mu_c = E[Y_c]$ and we can also identify μ_p, so the parameter $\widetilde{\Delta}_p = \mu_p - \mu_c$ is identified. ■

We will now state an assumption that produces identification of Δ_p, though it will only hold in special cases.

(A-7) $$E(Y_c \mid Z, d) = E(Y_c \mid d)$$

For some $z_1 \in Z$ and $z_2 \in Z$, $Pr(d = 1 \mid Z = z_1) \neq Pr(d = 1 \mid Z = z_2)$.

Note that this assumption only requires that the variable Z take on two distinct values. Much of the literature requires that the support Z be the whole real line $\Re^1$ or that there be some value $z \in Z$ for which $Pr(d = 1 \mid Z = z) = 1$. Neither of these is required for identification if we are willing to make assumption (A-7).

We first show that this assumption is sufficient for identification of Δ_p. Suppose assumption (A-7) holds for some particular values z_1 and z_2. We define $P_1 \equiv \Pr(d = 1 \mid Z = z_1)$ and $P_2 \equiv \Pr(d = 1 \mid Z = z_2)$ where $P_1 \neq P_2$. Then from the law of iterated expectations it follows that

$$E(Y_c|d = 0) = \frac{P_1 E(Y_c|Z = z_2) - P_2 E(Y_c|Z = z_1)}{P_1 - P_2}$$

and hence we can obtain $\Delta_p = E(Y_p \mid d = 0) - E(Y_c \mid d = 0)$, since $E(Y_p \mid d = 0)$ is known. This is analogous to the "average causal treatment effect" of Imbens and Angrist (1994) defined for the control state earnings of full program participants.

A major interpretive issue is whether assumption (A-7) is plausible. Except in very special cases, it is not. This assumption requires that Y_c be dependent on Z, but only through the seemingly irrelevant event corresponding to $d = 0$.

Assumption (A-7) will not--in general--be satisfied under conventional assumptions made in the discrete choice literature. In the model of example 1, if ε_c is dependent on ε_d and Z is stochastically independent of (ε_c, ν) then

$$E(Y_c|d = 0,Z) = E(Y_c|d = 0, Z\gamma + \nu < 0)$$

$$= \mu_c + \frac{\int_{-\infty}^{\infty}\int_{-\infty}^{-Z\gamma} \varepsilon_c f(\varepsilon_c,\nu)d\nu}{\int_{-\infty}^{-Z\gamma} f(\nu)\, d\nu}$$

where $f(\varepsilon_c, \nu)$ is the joint density of (ε, ν) and $f(\nu)$ is the density of ν. In this case, (A-7) is violated. Allowing Z to be stochastically dependent on ε_c, ν makes it possible for (A-7) to be satisfied through offsetting dependence effects.

In the following example, (A-7) is satisfied in a non-trivial way.

Example 2: Let

22(a) $\quad Z = \alpha_z\theta + U(z)$

22(b) $\quad Y_c = \alpha_Y\theta + U(Y_c)$

22(c) $\quad d = \theta$

where θ is a discrete random variable ($\theta = 0,1$) and (U(Z), U(Y)) are mutually independent and are also

independent of θ. Heuristically, persons with $\theta = 1$ are the unmotivated while those with $\theta = 0$ are motivated. Motivated persons do not drop out ($d = 0$ for $\theta = 0$). Let Z be ability. Then if $\alpha_Z < 0$, more motivated persons have higher ability. If $\alpha_Y < 0$, more motivated persons have higher income. In this example, it is clear that

$$E(Y_c \mid d, Z) = E(Y_c \mid d)$$

so (A-7) is satisfied. Dropout status is a perfect predictor of θ which drives the correlation among all three random variables.

If 22(c) is modified slightly to read

22(c)′ $$d = \theta + U(d)$$

where $E(U(d)) = 0$, U(d) is conditionally independent of θ, and U(Z), $U(Y_c)$, U(d) are mutually independent, then (A-7) is violated. Adding a little "noise" U(d) to θ in producing d, makes Z a useful predictor of Y_c even given d. In this setting, (A-7) is a very fragile assumption.

V. Conclusion

In this paper, we have examined several aspects of the Bloom (1984) estimator commonly used to produce estimates of the impact of treatment on the treated in the context of experimental evaluations in which not all treatment group members receive treatment. We have formalized the key assumption underlying this estimator, discussed the contexts in which it is likely to be empirically valid, and provided estimates using data from a recent experimental evaluation of the bias that results when it fails to hold. We show that the common practice of failing to account for the estimation of the dropout rate in estimating the variance of the estimator has little practical effect.

We generalize the Bloom estimator to estimate the full outcome distribution for the participant analogs in the control group and develop and apply statistical tests of the generalized form of Bloom's assumption based on restrictions on this constructed outcome distribution. We conclude by discussing alternative methods of obtaining identification of the mean impact of full treatment on the fully treated through the use of instruments or via alternative assumptions about earnings patterns within the treatment and control groups.

Underlying our examination of all of these aspects of the Bloom estimator are two important points. The first is that the Bloom estimator relies on an important assumption that may not hold in all evaluation contexts. The use of this estimator or one of the alternate methods of identification developed here must depend on the extent to which the assumptions invoked match the reality of a given evaluation

context. Inappropriate use of the Bloom estimator can result in substantial bias.

The second important point is that the mean impact of treatment on the treated as defined by the Bloom estimator is not always the parameter of economic interest. For example, in the case of programs with endowment or stigma effects, the mean impact of assignment to the treatment group is the parameter of interest, as everyone assigned to the treatment group is affected by the program, regardless of whether or not they participate in some formal sense. Clements, Heckman and Smith (1993) and Heckman (1992) discuss alternative parameters of interest.

Our analysis of the JTPA data reveals that the testable form of assumptions justifying Bloom's estimator are not rejected by the data. This evidence, the additional evidence that correcting for estimation of the proportion of persons who drop out and the evidence from our sensitivity analysis appear to vindicate Bloom's estimator in its application to the JTPA data.

Bibliography

Amemiya, Takeshi. 1985. Advanced Econometrics, Cambridge, MA., Harvard University Press.

Angrist, Joshua and Guido Imbens. 1991. "Sources of Identifying Information in Evaluation Models." Unpublished Manuscript, Harvard University, Cambridge, Massachusetts.

Bloom, Howard. 1984. "Accounting for No-shows in Experimental Evaluation Designs." *Evaluation Review* 8(2), 225-246.

Bloom, Howard, Larry Orr, George Cave, Steve Bell and Fred Doolittle. 1993. The National JTPA Study: Title IIA Impacts on Earnings and Employment at 18 Months. Bethesda, MD: Abt Associates.

Burtless, Gary. 1985. "Are Targeted Wage Subsidies Harmful? Evidence From A Wage Voucher Experience." *Industrial and Labor Relations Review* 39, 105-114.

Clements, Nancy, J. Heckman and J. Smith. 1993. "Making the Most Out of Randomized Social Experiments", unpublished manuscript, University of Chicago, March, 1993, revised November, 1993.

Dubin, Jeffrey and Douglas Rivers. 1993. "Experimental Estimates of the Impact of Wage Subsidies." *Journal of Econometrics* 56, 219-242.

Efron, Bradley and David Feldman. 1991. "Compliance As An Explanatory Variable in Clinical Trials", Journal of The American Statistical Association, 8, 9-17.

Gueron, Judith and Mark Pauly. 1991. From Work to Welfare. New York: Russell Sage Foundation.

Haynes, Brian and Renato Dantes. 1987. "Patient Compliance and the Conduct and Interpretation of Therapeutic Trials." *Controlled Clinical Trials* 8(1), 12-19.

Heckman, James. 1990. "Varieties of Selection Bias." *American Economic Review* 80.

Heckman, James. 1992. "Randomization and Social Policy Evaluation." in Charles Manski and Irwin Garfinkel, eds., Evaluating Welfare and Training Programs. Cambridge, MA: Harvard University Press.

Heckman, James and Jeffrey Smith. 1994. "Assessing the Case For Randomized Evaluation of Social Programs." in The Effects and Measuring the Effects of Labour Market Policy Initiatives. Copenhagen: Danish Ministry of Labor.

Hotz, V. J. and Seth Sanders. 1994. "Bounding Treatment Effects in Controlled and Natural Experiments Subject to Post-Randomization Treatment Choice", working paper, Population Research Center, 94-2.

Imbens, Guido and Joshua Angrist. 1994. "Identification and Estimation of Local Average Treatment Effects." *Econometrica* 62(2), 467-476.

Kemple, James, Fred Doolittle and Neil Wallace. 1993. The National JTPA Study: Site Characteristics and Participation Patterns. New York, Manpower Demonstration Research Corporation.

Table 1

RATIO OF ESTIMATED VARIANCES OF THE IMPACT PER ENROLLER
CORRECTED AND NOT CORRECTED FOR THE ESTIMATION OF THE DROPOUT RATE

JTPA EXPERIMENTAL DATA FOR ADULT MALES

SUM OF EARNINGS IN THE 18 MONTHS AFTER RANDOM ASSIGNMENT

$E(Y_c)-E(Y_t \mid d=1)$	k = 0.01	k = 0.05	k = 0.10	k = 0.25	k = 0.50
\$0.00	1.00000	1.00000	1.00000	1.00000	1.00000
\$250.00	1.00000	1.00001	1.00002	1.00005	1.00016
\$500.00	1.00001	1.00003	1.00007	1.00021	1.00062
\$1000.00	1.00003	1.00013	1.00028	1.00083	1.00249
\$2000.00	1.00010	1.00053	1.00112	1.00334	1.00995

1. Table updated on May 29, 1994
2. The control group is assumed to be one half the size of the treatment group, as in the National JTPA study.
3. k is the fraction of the treatment group dropping out of the program after random assignment. $E(Y_c)$ is the earnings mean of the controls while $E(Y_t \mid d=1)$ is the earnings mean of the treatment group dropouts.
4. The calculations use the estimated variances of earnings in the 18 months after random assignment for adult males in the National JTPA Study. The corresponding estimated standard deviations are \$11,593.97 for controls, \$11,804.85 for treatment group dropouts and \$11,542.70 for treatment group participants.
5. Note that the ratio of earnings variances does not depend on the sample size.

TABLE 2

DIFFERENCE IN ESTIMATED STANDARD ERRORS OF THE IMPACT PER ENROLLEE CORRECTED AND NOT CORRECTED FOR THE ESTIMATION OF THE DROPOUT RATE

JTPA EXPERIMENTAL DATA FOR ADULT MALES

SUM OF EARNINGS IN THE 18 MONTHS AFTER RANDOM ASSIGNMENT

# in Treatment Group	k = 0.01	k = 0.05	k = 0.10	k = 0.25	k = 0.50
$E(Y_c) - E(Y_t \mid d = 1) = \$0.00$					
300	0.0000	0.0000	0.0000	0.0000	0.0000
600	0.0000	0.0000	0.0000	0.0000	0.0000
1200	0.0000	0.0000	0.0000	0.0000	0.0000
2100	0.0000	0.0000	0.0000	0.0000	0.0000
3000	0.0000	0.0000	0.0000	0.0000	0.0000
$E(Y_c) - E(Y_t \mid d = 1) = \$250.00$					
300	0.0010	0.0050	0.0111	0.0402	0.1799
600	0.0006	0.0035	0.0079	0.0284	0.1272
1200	0.0005	0.0025	0.0056	0.0201	0.0900
2100	0.0003	0.0019	0.0042	0.0151	0.0680
3000	0.0003	0.0016	0.0035	0.0127	0.0569
$E(Y_c) - E(Y_t \mid d = 1) = \$500.00$					
300	0.0038	0.0200	0.0447	0.1605	0.7195
600	0.0026	0.0142	0.0316	0.1135	0.5089
1200	0.0019	0.0100	0.0223	0.0803	0.3597
2100	0.0014	0.0076	0.0169	0.0607	0.2720
3000	0.0012	0.0063	0.0142	0.0508	0.2275
$E(Y_c) - E(Y_t \mid d = 1) = \$1000.00$					
300	0.0148	0.0802	0.1787	0.6420	2.8770
600	0.0104	0.0568	0.1264	0.4540	2.0344
1200	0.0074	0.0401	0.0894	0.3210	1.4385
2100	0.0056	0.0303	0.0676	0.2426	1.0874
3000	0.0047	0.0254	0.0565	0.2030	0.9098
$E(Y_c) - E(Y_t \mid d = 1) = \$2000.00$					
300	0.0592	0.3210	0.7148	2.5663	11.4868
600	0.0418	0.2271	0.5055	1.8147	8.1224
1200	0.0296	0.1605	0.3574	1.2831	5.7434
2100	0.0223	0.1213	0.2702	0.9700	4.3416
3000	0.0187	0.1015	0.2261	0.8116	3.6324

1. Table updated on May 29, 1994
2. The control group is assumed to be one half the size of the treatment group, as in the National JTPA study.
3. k is the fraction of the treatment group dropping out of the program after random assignment. $E(Y_c)$ is the earnings mean of the controls while $E(Y_t \mid d = 1)$ is the earnings mean of the treatment group dropouts.
4. The calculations use the estimated variances of earnings in the 18 months after random assignment for adult males in the National JTPA Study. The corresponding estimated standard deviations are \$11,593.97 for controls, \$11,804.85 for treatment group dropouts and \$11,542.70 for treatment group participants.

TABLE 3

SAMPLE MEAN EARNINGS, k, AND MEAN DIFFERENCE IMPACT ESTIMATES
Full Abt 18 Month Impact Sample

Target Group Outcome	Adult Men	Adult Women	Male Youth	Female Youth
$E(Y_t)$ - treatment group	13096.43 (210.78)	8261.81 (132.31)	9997.76 (253.98)	6163.67 (163.33)
Full participants $E(Y_t \mid d = 0) = E(Y_p \mid d = 0)$	13638.23 (260.96)	8424.97 (155.89)	10274.77 (310.36)	6114.49 (196.76)
Less than full participation $E(Y_t \mid d = 0)$	12181.06 (354.80)	7946.47 (244.35)	9442.33 (441.13)	6256.13 (290.81)
$E(Y_c)$ - controls	12530.09 (305.57)	7470.98 (180.20)	10781.72 (401.80)	6202.09 (248.77)
k - fraction dropping out	0.3718	0.3410	0.3328	0.3472
Δ - impact estimate	566.34 (371.21)	790.83 (223.56)	-783.96 (475.34)	-38.42 (297.60)
N - sample size	4420	5724	1747	2301

Source: Self-reported earnings data from the National JTPA Study.

NOTES

1. Adult female non-respondents not included.

2. Estimated standard errors in parentheses.

TABLE 4

PERCENTAGE DISTRIBUTION OF
POST-RANDOM ASSIGNMENT ACTIVITY IN JTPA
OF TREATMENT GROUP MEMBERS WHO DID NOT ENROLL

Activity	Nonenrollees (%)
No Further Contact	15
Further Contact, But Not Eligible	1
No Longer Interested (a)	11
Get job on own	5
Moved	2
Health problems	1
In another program	1
Reason unknown	3
Interested, But Made Contact Only and Received No Services	20
Interested and Received Service(s) (b)	53
Received further assessment and counseling	11
Referred to classroom training provider(s)	5
Received support service(s)	2
Referred to employer(s) for possible on-the-job training	36
Participated in job club or received job search assistance	20
Total	100
Sample Size	307

Source: Information collected by MDRC site representatives during the National JTPA Study, presented as Table 3.2 in Kemple, Doolittle and Wallace (1993).

NOTES: Calculations for this table are based on data for a random sample of 307 treatment group members in the 18-month study sample who did not enroll in JTPA.

(a) When totaled, the subcategory percentages are over 11 percent because nonenrollees could cite more than one reason for no longer being interested in JTPA.

(b)When totaled, the subcategory percentages are over 53 percent because some nonenrollees received more than one service.

TABLE 5

ESTIMATED IMPACTS ON EARNINGS IN THE 18 MONTHS AFTER RANDOM ASSIGNMENT
Full Abt 18 Month Impact Sample

Target Group	Adult Men	Adult Women	Male Youth	Female Youth
Bloom estimate Δ_p	901.55 (590.60)	1200.00 (339.19)	-1174.96 (712.30)	-58.85 (455.92)
Δ^{ϵ}_p with ϵ = 150	990.33 [88.78]	1277.61 [77.61]	-1100.15 [74.81]	20.92 [79.78]
Δ^{ϵ}_p with ϵ = 100	960.74 [59.19]	1251.74 [51.74]	-1125.08 [49.87]	-5.67 [53.19]
Δ^{ϵ}_p with ϵ = 50	931.14 [29.59]	1225.87 [25.87]	-1150.02 [24.94]	-32.26 [26.59]
Δ^{ϵ}_p with ϵ = -50	871.95 [-29.59]	1174.13 [-25.87]	-1199.90 [-24.94]	-85.45 [-26.59]
Δ^{ϵ}_p with ϵ = -100	842.36 [-59.19]	1148.26 [-51.74]	-1224.83 [-49.87]	-112.04 [-53.19]
Δ^{ϵ}_p with ϵ = -150	812.77 [-88.78]	1122.39 [-77.61]	-1249.77 [-74.81]	-138.63 [-79.78]

Source: Self-reported earnings data from the National JTPA Study.

NOTES

1. The estimates presented here differ from those in Bloom et al. (1993) because (1) estimates are calculated using simple means without regression adjustment and (2) imputed values for adult female non-respondents based on UI earnings data were not used.

2. $\Delta_p = E(Y_t \mid d=1) - E(Y_d \mid d=1) = E(Y_t \mid d=1) - [1/(1-k)]\ E(Y_c) + [k/(1-k)]\ E(Y_t \mid d=1)$

3. $\Delta^{\epsilon}_p = (Y_t \mid d=1) - E(Y_c \mid d=1) = E(Y_t \mid d=1) - [1/(1-k)]\ E(Y_c) + [k/(1-k)]\ E(Y_t \mid d=1) + [k/(1-k)]\epsilon$

4. Estimated standard errors in parentheses. Since ϵ is a constant, the standard errors for Δ^{ϵ}_p equal those for Δ_p and are not repeated.

5. The estimated bias = $\Delta^{\epsilon}_p - \Delta_p$ appears in square brackets.

TABLE 6

ESTIMATED IMPACTS ON EARNINGS IN THE 18 MONTHS AFTER RANDOM ASSIGNMENT
BASED ON ASSUMPTION (A-5): $E(Y_c \mid d=1) = \eta E(Y_c \mid d=0)$
Full Abt 18 Month Impact Sample

Target Group	Adult Men	Adult Women	Male Youth	Female Youth
Bloom estimate Δ_p	901.55 (590.60)	1200.00 (339.19)	-1174.96 (712.30)	-58.85 (455.92)
Δ^{η}_p with $\eta = 0.50$	-1753.23 (457.15)	-581.53 (267.38)	-2658.96 (573.28)	-1390.44 (359.63)
Δ^{η}_p with $\eta = 0.75$	-175.93 (426.13)	257.77 (251.21)	-1485.31 (537.03)	-677.10 (336.04)
Δ^{η}_p with $\eta = 1.00$	1108.13 (401.84)	953.99 (238.27)	-506.94 (507.71)	-87.60 (317.18)
Δ^{η}_p with $\eta = 1.25$	2173.79 (382.45)	1540.82 (227.76)	321.14 (483.65)	407.74 (301.85)
Δ^{η}_p with $\eta = 1.50$	3072.39 (366.73)	2042.18 (219.10)	1031.09 (463.67)	829.81 (289.22)
Value of η such that $\Delta^{\eta}_p = \Delta_p$	0.9564	1.0999	0.8247	1.0134

Source: Self-reported earnings data from the National JTPA Study.

NOTES

1.The estimates presented here differ from those in Bloom et al. (1993) because (1) estimates are calculated using simple means without regression adjustment and (2) imputed values for adult female non-respondents based on UI earnings data were not used.

2.$\Delta_p = E(Y_t \mid d=1) - E(Y_c \mid d=1) = E(Y_t \mid d=1) - [1/(1-k)]\ E(Y_c) + [k/(1-k)]\ E(Y_t \mid d=1)$

3.$\Delta^{\eta}_p = E(Y_p) - Y^{*}_p = E(Y_t \mid d=1) - \{E(Y_c) / [1 + k(\eta-1)]\}$

4.Estimated standard errors in parentheses.

APPENDIX

Least Favorable Distributions

In this appendix we develop a test of restriction (R-1) and discuss theoretical and Monte Carlo results on the least favorable distribution for the test statistic. For the null hypothesis there is no similar test statistic; the size of the test will vary among different distributions which are consistent with it. In this situation, the size of the test must be defined as the maximum probability of rejecting the null hypothesis. A least favorable distribution is a distribution for which the probability of rejection equals this maximum. In general there will be more than one such null. The goal of this appendix is to characterize the attributes of some least favorable distributions, though not necessarily all least favorable distributions.

For simplicity, we begin with a model with three cells that partition the support of Y_c. Let $P_c(i)$ denote the probability that Y_c lies in cell b_i, $P_d(i)$ denote the probability that $(Y_t \mid d = 1)$ lies in cell b_i, and let k denote the fraction of dropouts, which is assumed to be fixed. Let $\hat{P}_c(i)$ and $\hat{P}_d(i)$, be their sample analogs. We define

$$P_p^*(i) = \frac{P_c(i) - kP_d(i)}{1-k}$$

with estimator

$$\hat{P}_p^*(i) = \frac{\hat{P}_c(i) - k\hat{P}_d(i)}{1-k} \quad .$$

Furthermore let

$$P_c = \begin{bmatrix} P_c(1) \\ P_c(2) \\ P_c(3) \end{bmatrix} \quad P_d = \begin{bmatrix} P_d(1) \\ P_d(2) \\ P_d(3) \end{bmatrix} \quad P_p^* = \begin{bmatrix} P_p^*(1) \\ P_p^*(2) \\ P_p^*(3) \end{bmatrix}$$

and

$$\hat{P}_c = \begin{bmatrix} \hat{P}_c(1) \\ \hat{P}_c(2) \\ \hat{P}_c(3) \end{bmatrix} \quad \hat{P}_d = \begin{bmatrix} \hat{P}_d(1) \\ \hat{P}_d(2) \\ \hat{P}_d(3) \end{bmatrix} \quad \hat{P}_p^* = \begin{bmatrix} \hat{P}_p^*(1) \\ \hat{P}_p^*(2) \\ \hat{P}_p^*(3) \end{bmatrix}.$$

It is easy to show that

$$\sqrt{n}(P_c - \hat{P}_c) \rightarrow N(0,\Sigma_c)$$

$$\sqrt{n}(P_d - \hat{P}_d) \rightarrow N(0,\Sigma_d)$$

$$\sqrt{n}(P_p^* - \hat{P}_p^*) \rightarrow N\left(0,\frac{\Sigma_c + k^2\Sigma_d}{(1-k)^2}\right)$$

where

$$\Sigma_c = \begin{bmatrix} P_c(1)(1-P_c(1)) & -P_c(1)P_c(2) & -P_c(1)P_c(3) \\ -P_c(1)P_c(2) & P_c(2)(1-P_c(2)) & -P_c(2)P_c(3) \\ -P_c(1)P_c(3) & -P_c(2)-P_c(3) & P_c(3)(1-P_c(3)) \end{bmatrix}$$

and

$$\Sigma_d = \begin{bmatrix} P_d(1)(1-P_d(1)) & -P_d(1)P_d(2) & -P_d(1)P_d(3) \\ -P_d(1)P_d(2) & P_d(2)(1-P_d(2)) & -P_d(2)P_d(3) \\ -P_d(1)P_d(3) & -P_d(2)-P_d(3) & P_d(3)(1-P_d(3)) \end{bmatrix}.$$

We consider the following null hypothesis:

$$H_0: \quad \begin{aligned} P_p^*(1) &\geq 0 \\ P_p^*(2) &\geq 0 \\ P_p^*(3) &\geq 0. \end{aligned}$$

If our goal was to test a single inequality restriction, we would create the t-statistic and perform a t-test. In a similar manner, with more than one inequality constraint, we can create

a vector of t-statistics. Let t denote this vector and t(i) denote the ith component of the vector. For each individual t-statistic

$$t(i) \rightarrow \begin{cases} \infty & P_p^*(i) > 0 \\ N(0,1) & P_p^*(i) = 0 \\ -\infty & P_p^*(i) < 0 \end{cases}$$

One possible generalization of the t-test is to define a critical region by the vector c = (c(1), c(2), c(3))′ and reject if t(i) < c(i) for any i = {1, 2, 3}. For a related test statistic, Wolak (1991) showed that the constraint will bind for at least two of the cells at the least favorable distribution. We will show this to be true for our test statistic as well.

An attractive feature of our specification is that a priori there is no difference between any of the cells. Furthermore,

$$P_p^*(1) + P_p^*(2) + P_p^*(3) = 1,$$

so the constraint can bind in at most two cells. Without loss of generality we will ignore the case $P_p^*(3) = 0$. Let $\Phi(c)$ denote the cdf of a standard normal and let $\Phi(c_1, c_2; \rho)$ denote the cdf of a two dimensional normal with unit variances and correlation coefficient ρ. It is easy to show that the probability of accepting the null hypothesis can be written as follows;

$$\lim_{n\to\infty} Pr(t > c) = \begin{cases} 1.00 & P_p^*(1) > 0, P_p^*(2) > 0, P_p^*(3) > 0 \\ \Phi(-c(1)) & P_p^*(1) = 0, P_p^*(2) > 0, P_p^*(3) > 0 \\ \Phi(-c(1), -c(2); \rho_{12}) & P_p^*(1) = 0, P_p^*(2) = 0, P_p^*(3) > 0 \end{cases}$$

where ρ_{12} is the asymptotic correlation between t(1) and t(2).

A convenient property of the joint normal cdf is that given c(1) and c(2), $\Phi(-c(1), -c(2); \rho_{12})$ is a strictly increasing function of ρ_{12}. Using this fact, it is easy to show that a least favorable distribution will occur when two constraints are satisfied. Without loss of generality assume $-c(1) \leq -c(2)$. First notice that when $\rho_{12} = 1$,

$$\Phi(-c(1), -c(2); 1) = \Phi(-c(1)).$$

This implies that for general ρ_{12},

$$\Phi(-c(1), -c(2); \rho_{12}) \leq \Phi(-c(1)),$$

so that the constraint will bind in two cells.

A least favorable distribution can be derived by minimizing the correlation with respect to $(P_c(1), P_c(2), P_d(1), P_d(2))$ subject to the constraints

$$P_c(1) = k\, P_d(1)$$

$$P_c(2) = k\, P_d(2)$$

$$1 - P_d(1) - P_d(2) \geq 0.$$

The solution to this problem is the following:

$$P_c = \begin{bmatrix} \frac{k}{2} \\ \frac{k}{2} \\ 1-k \end{bmatrix} \quad P_d = \begin{bmatrix} \frac{1}{2} \\ \frac{1}{2} \\ 0 \end{bmatrix}.$$

At this solution $\rho_{12} = -k$.

In practice, the size of the test will not be very sensitive to ρ_{12} as long as ρ_{12} is negative. Suppose that $c = c(1) = c(2)$ and $\Phi(-c) = 1 - \alpha$. Consider the following two extreme cases

$$\Phi(-c, -c; -1) = \Pr(\,|\,t\,|\, > c) = 1 - 2\alpha$$

$$\Phi(-c, -c; 0) = \Pr(t(1) > c)\, \Pr(t(2) > c) = (1 - \alpha)^2$$

However, since $(1 - \alpha)^2 - (1 - 2\alpha) = \alpha^2$ and α is small even at the two extremes, the difference in the size of the test will be very small. For example when $\alpha = 0.025$, $1 - 2\alpha = 0.95$ and $(1 - \alpha)^2 = 0.9506$.

We performed Monte Carlo study to examine the sensitivity of the size of the test to various distribution of the data consistent with the null hypothesis. We used 10,000 Monte Carlo draws each with sample size 2000 and $k = ½$. In Table A1 we present the fraction of Monte Carlo draws for which the null hypothesis was accepted. The predictions of the theoretical discussion are borne out in the Monte Carlo results.

All three critical regions were chosen so that asymptotically the size of the test should be close to .05. Since the constraint doesn't bind in the third cell for any of the true distributions considered, asymptotically only c(1) and c(2) are relevant. The relative sizes of c(1) and c(2) are chosen to be the same in the first column. We predict that the probability of rejecting is .95 when the first two constraints bind, .975 when only one binds, and 1.0 when none bind. The second critical region is not symmetric, the probability or rejecting based on c(1) is .964, the probability of rejecting based on c(2) is .986, and the joint probability of rejecting is approximately .95 when both constraints bind. The third critical region was chosen so that the probability of rejecting based on c(2) is very small, so the probability of rejecting based on c(1) is .95 and the joint probability is .95. These theoretical predictions are very close to the Monte Carlo predictions. It appears that 2000 observations is sufficient to invoke asymptotic theroy.

We also performed some Monte Carlo runs to gauge the power of the tests. These results appear in Table A2. Note that it is the probability of accepting the null hypothesis that is reported rather than the power itself. The test is reasonably powerful against moderately - sized violations of the null hypothesis that are concentrated in a particular cell.

Rather than testing the individual cells, we may want to test whether the cdf remains bounded between zero and one. Using three cells as before, we can write this restriction in terms of the probabilities defined above:

$$H_0: \quad \begin{aligned} P_p^*(1) &\geq 0 \\ P_p^*(1) + P_p^*(2) &\geq 0 \\ P_p^*(2) + P_p^*(3) &\geq 0 \\ P_p^*(3) &\geq 0. \end{aligned}$$

Note that the condition

$$P_p^*(1) + P_p^*(2) + P_p^*(3) = 1,$$

is automatically imposed. We can proceed exactly as before. We obtain the vector of test statistics by forming the sample analogs of the four conditions above and dividing each by its standard error. The test is analogous to the one performed above.

Note that the conditions imposed here are weaker than before. It is possible for $P_p^*(2)$ to be negative but still satisfy this null hypothesis. However, if the previous conditions hold then

these conditions must hold as well.

As before the constraint can only bind in at most two of the cells, so we can appeal to our earlier argument. The only difference is that the correlations of the t-statistics will be somewhat different than before. The least favorable distribution will occur when the constraint binds in two of the cells. Notice that if the two constraints that bind are $P_p^*(1) = 0$ and $P_p^*(3) = 0$, then the correlation will be negative. For any other two constraints, the correlation will be positive. We know the correlation is minimized at the least favorable distribution, so a least favorable distribution must occur when $P_p^*(1) = 0$ and $P_p^*(3) = 0$. In this case the least favorable distribution will be exactly the same as the one derived above.

We present Monte Carlo results for these test statistics in Tables A3 and A4. Note that since the case $P_p^*(1) = 0$ and $P_p^*(3) = 0$ is analogous to the case $P_p^*(1) = 0$ and $P_p^*(2) = 0$ previously reported, we don't report it here.

As in our previous results, the predicted probabilities are very close to their Monte Carlo analogues. For all three test statistics the least favorable distribution wil occur when $P_p^*(1) = 0$ and $P_p^*(3) = 0$ where the probability of acceptance is .95. The advantage of the cdf test is that it should have more power against some alternatives. This result is borne out in Table A4 where the cdf test has more power against violations of the null hypothesis in both the first and second cell.

We now return to the original test but allow an arbitrary number of cells. In this case it is trivial to show that the constraint will bind in K - 1 of the cells for at least one least favorable distribution.

Suppose this was not the case. Suppose that the constraint binds in only $K^* < K - 1$ of the cells. Asymptotically the probability of accepting the null hypothesis depends only on the K^* cells for which the constraint binds. Since the cell probabilities in the other $K - K^*$ are irrelevant we can change them arbitrarily without affecting the probability of accepting the null hypothesis. Therefore we can redefine a new least favorable distribution by taking all of the mass from the $K - K^*$ cells for which the constraint does not bind and putting it into a single one of those cells. The probability of accepting remains unchanged, but the constraint now binds in K - 1 of the cells. This new null hypothesis may be uninteresting, but it does demonstrate that in searching for the least favorable null hypothesis we can restrict ourselves to ones in which the constraint

binds in K - 1 of the cells.

TABLE A1

Monte Carlo Evidence on Least Favorable Distribution

Probability of Acceptance [a]

True Distribution	Critical Region		
	$c(1) = -1.96$	$c(1) = -1.80$	$c(1) = -1.65$
	$c(2) = -1.96$	$c(2) = -2.20$	$c(2) = -3.30$
	$c(3) = -1.96$	$c(3) = -1.80$	$c(3) = -1.65$
$P_c = (0.25, 0.25, 0.50)$ $P_d = (0.50, 0.50, 0.00)$ $P_p = (0.00, 0.00, 1.00)$	0.950	0.948	0.948
$P_c = (0.20, 0.20, 0.60)$ $P_d = (0.40, 0.40, 0.20)$ $P_p = (0.00, 0.00, 1.00)$	0.946	0.946	0.947
$P_c = (0.30, 0.10, 0.60)$ $P_d = (0.60, 0.20, 0.20)$ $P_p = (0.00, 0.00, 1.00)$	0.947	0.945	0.949
$P_c = (0.25, 0.26, 0.49)$ $P_d = (0.50, 0.50, 0.00)$ $P_p = (0.00, 0.02, 0.98)$	0.971	0.960	0.945
$P_c = (0.25, 0.30, 0.45)$ $P_d = (0.50, 0.50, 0.00)$ $P_p = (0.00, 0.10, 0.90)$	0.976	0.965	0.951
$P_c = (0.25, 0.25, 0.50)$ $P_d = (0.50, 0.49, 0.01)$ $P_p = (0.00, 0.01, 0.99)$	0.964	0.957	0.948
$P_c = (0.25, 0.25, 0.50)$ $P_d = (0.50, 0.25, 0.25)$ $P_p = (0.00, 0.25, 0.75)$	0.973	0.961	0.947
$P_c = (0.26, 0.26, 0.48)$ $P_d = (0.50, 0.50, 0.00)$ $P_p = (0.02, 0.02, 0.96)$	0.994	0.995	0.993
$P_c = (0.30, 0.30, 0.40)$ $P_d = (0.50, 0.50, 0.00)$ $P_p = (0.10, 0.10, 0.80)$	1.000	1.000	1.000

[a]These results were obtained using 10,000 Monte Carlo runs, with sample size 2000, and k=0.5

TABLE A2

Monte Carlo Evidence on Power of Test

Probability of Acceptance[a]

True Distribution	Critical Region		
	$c(1) = -1.96$ $c(2) = -1.96$ $c(3) = -1.96$	$c(1) = -1.80$ $c(2) = -2.20$ $c(3) = -1.80$	$c(1) = -1.65$ $c(2) = -3.30$ $c(3) = -1.65$
$P_c = (0.24, 0.25, 0.51)$ $P_d = (0.50, 0.50, 0.00)$ $P_p = (-0.02, 0.00, 1.02)$	0.822	0.793	0.764
$P_c = (0.24, 0.38, .38)$ $P_d = (0.50, 0.25, 0.25)$ $P_p = (-0.02, 0.51, 0.51)$	0.851	0.805	0.765
$P_c = (0.24, 0.24, 0.52)$ $P_d = (0.50, 0.50, 0.00)$ $P_p = (-0.02, -0.02, 1.04)$	0.851	0.805	0.765
$P_c = (0.33, 0.33, 0.34)$ $P_d = (0.67, 0.17, 0.16)$ $P_p = (-0.01, 0.49, 0.52)$	0.932	0.909	0.882
$P_c = (0.30, 0.35, 0.35)$ $P_d = (0.67, 0.17, 0.16)$ $P_p = (-0.07, 0.53, 0.54)$	0.144	0.112	0.087

[a]These results were obtained using 10,000 Monte Carlo runs, with sample size 2000, and k=0.5

TABLE A3

Monte Carlo Evidence on Least Favorable Distribution

CDF Test

Probability of Acceptance [a]

True Distribution	Critical Region		
	$c(1) = -1.96$ $c(2) = -1.96$ $c(3) = -1.96$ $c(4) = -1.96$	$c(1) = -1.80$ $c(2) = -2.20$ $c(3) = -1.80$ $c(4) = -1.80$	$c(1) = -1.65$ $c(2) = -3.30$ $c(3) = -1.65$ $c(4) = -1.65$
$P_c = (0.25, 0.25, 0.50)$ $P_d = (0.50, 0.50, 0.00)$ $P_p = (0.00, 0.00, 1.00)$	0.969	0.961	0.949
$P_c = (0.20, 0.20, 0.60)$ $P_d = (0.40, 0.40, 0.20)$ $P_p = (0.00, 0.00, 1.00)$	0.963	0.957	0.948
$P_c = (0.30, 0.10, 0.60)$ $P_d = (0.60, 0.20, 0.20)$ $P_p = (0.00, 0.00, 1.00)$	0.965	0.959	0.950
$P_c = (0.25, 0.26, 0.49)$ $P_d = (0.50, 0.50, 0.00)$ $P_p = (0.00, 0.02, 0.98)$	0.973	0.960	0.945
$P_c = (0.25, 0.30, 0.45)$ $P_d = (0.50, 0.50, 0.00)$ $P_p = (0.00, 0.10, 0.90)$	0.976	0.965	0.951
$P_c = (0.25, 0.25, 0.50)$ $P_d = (0.50, 0.49, 0.01)$ $P_p = (0.00, 0.01, 0.99)$	0.970	0.961	0.948
$P_c = (0.25, 0.25, 0.50)$ $P_d = (0.50, 0.25, 0.25)$ $P_p = (0.00, 0.25, 0.75)$	0.973	0.961	0.947
$P_c = (0.26, 0.26, 0.48)$ $P_d = (0.50, 0.50, 0.00)$ $P_p = (0.02, 0.02, 0.96)$	0.998	0.997	0.994
$P_c = (0.30, 0.30, 0.40)$ $P_d = (0.50, 0.50, 0.00)$ $P_p = (0.10, 0.10, 0.80)$	1.000	1.000	1.000

[a] These results were obtained using 10,000 Monte Carlo runs, with sample size 2000, and k=0.5

TABLE A4

Monte Carlo Evidence on Power of Test

CDF Test

Probability of Acceptance [a]

True Distribution	Critical Region		
	$c(1) = -1.96$ $c(2) = -1.96$ $c(3) = -1.96$ $c(4) = -1.96$	$c(1) = -1.80$ $c(2) = -2.20$ $c(3) = -1.80$ $c(4) = -1.80$	$c(1) = -1.65$ $c(2) = -3.30$ $c(3) = -1.65$ $c(4) = -1.65$
$P_c = (0.24, 0.25, 0.51)$ $P_d = (0.50, 0.50, 0.00)$ $P_p = (-0.02, 0.00, 1.02)$	0.814	0.794	0.770
$P_c = (0.24, 0.38, .38)$ $P_d = (0.50, 0.25, 0.25)$ $P_p = (-0.02, 0.51, 0.51)$	0.850	0.811	0.766
$P_c = (0.24, 0.24, 0.52)$ $P_d = (0.50, 0.50, 0.00)$ $P_p = (-0.02, -0.02, 1.04)$	0.662	0.700	0.765
$P_c = (0.33, 0.33, 0.34)$ $P_d = (0.67, 0.17, 0.16)$ $P_p = (-0.01, 0.49, 0.52)$	0.936	0.914	0.890
$P_c = (0.30, 0.35, 0.35)$ $P_d = (0.67, 0.17, 0.16)$ $P_p = (-0.07, 0.53, 0.54)$	0.145	0.111	0.085

[a]These results were obtained using 10,000 Monte Carlo runs, with sample size 2000, and k=0.5