Measuring the Effect of Subsidized Training Programs on Movements In and Out of Employment

David Card; Daniel Sullivan

Stable URL:

Econometrica is currently published by The Econometric Society.

MEASURING THE EFFECT OF SUBSIDIZED TRAINING PROGRAMS ON MOVEMENTS IN AND OUT OF EMPLOYMENT

By David Card and Daniel Sullivan[1]

Despite over two decades of U.S. experience in operating large scale subsidized training programs for low income and unemployed workers, the effects of these programs are still highly controversial. The controversy arises from the difficulty of specifying the model of participant outcomes in the absence of training that is necessary in any nonexperimental program evaluation. In this paper we suggest that some of these difficulties may be overcome by focusing on a very simple measure of outcomes: namely, the probability of employment. We present a variety of estimates of the effect of training on the probability of employment for the 1976 cohort of adult male participants in the Comprehensive Employment and Training Act (CETA) program. Our methods range from a simple comparison of pre- and post-training employment probabilities between trainees and nonparticipants, to a fully specified first-order Markov model of employment probabilities with individual heterogeneity. There is consistent evidence across methods that CETA participation increased the probability of employment in the three years after training by from 2 to 5 percentage points. Classroom training programs appear to have had significantly larger effects than on-the-job programs, although the estimated effects of both kinds of programs are positive. We also find that movements in and out of employment for trainees and nonparticipants are reasonably well described by a first-order process, conditional on individual heterogeneity. In the context of this model, CETA participation appears to have increased both the probability of moving into employment, and the probability of continuing employment.

KEYWORDS: Program evaluation, selection bias, employment probability, Markov model, individual heterogeneity.

DURING THE PAST TWO DECADES the U.S. government has sponsored a series of large-scale subsidized training programs for unemployed and low-income workers.[2] The precise impact of these programs, however, remains a source of continuing controversy. At issue are the effects of training on the earnings of participants. The measurement of training effects in the absence of classical random assignment into treatment and control groups has proved exceedingly difficult, in part because of the difficulty of modelling the process of selection into training, and in part because of the difficulty of specifying a model of earnings in the absence of training or selection effects.[3]

[2] These programs were initiated by the Manpower Development and Training Act of 1962, modified and expanded by the Comprehensive Employment and Training Act of 1973, and recently restructured by the Job Partnership Training Act of 1982.

[3] Many of these difficulties were pointed out by Ashenfelter (1975). The reliability of various econometric techniques for program evaluation has been studied by LaLonde (1986), who applies nonexperimental estimators to data from the National Supported Work Demonstration. Comparing the nonexperimental program estimates to the experimentally determined training effect, LaLonde finds that the nonexperimental methods are sensitive to specification, and that conventional specification tests do not always provide a clear basis to choose between the diverse nonexperimental estimates.

For any training program, the effect on participant earnings can be decomposed into an effect on the probability of employment, and an effect on the level of earnings, conditional on employment. In this paper, we provide estimates of the former effect for adult male participants in the 1976 Comprehensive Employment and Training Act (CETA) programs. Our motivation for focussing on the employment probabilities is threefold. First, the data that we employ are nonexperimental. In a nonexperimental setting it is imperative to have an adequate model of the process generating the data. Past attempts to directly specify models for the level of earnings, however, have not been entirely successful. When, as in Ashenfelter and Card (1985), goodness-of-fit tests are available, the models are usually rejected. Moreover, the administrative earnings records that form the basis for our analysis exhibit a variety of features not easily incorporated in the simple variance components and time series models used in the literature. Many individuals in our sample report spells of zero earnings, while the earnings records for a significant fraction of the sample are censored. Indeed, it is our belief that the simplest way to specify an acceptable model of these data may be to combine the kind of employment probability models presented in this paper with some relatively simple model for the level of earnings conditional on employment.[4]

Secondly, we believe that it may be easier to model the effects of training on the employment probabilities of the trainees than on the level of their earnings. We focus on two distinct training effects on post-training employment probabilities. The first is an effect on the labor market status of trainees at the completion of training. To the extent that subsequent employment probabilities are affected by this status, training provides a one-time shift in the pattern of post-training employment probabilities. The second is an effect of training on the probability that trainees remain employed, or move from unemployment to employment. By modelling these two effects separately, we can provide a more complete description of the overall training effect.

Finally, there is evidence from a randomized trial evaluation of one recent training program that most of the measured training effect in that program resulted from increases in the post-training employment rates of the trainees.[5] Thus our analysis of employment probabilities for CETA trainees can be expected to capture a significant fraction of the total effect of training.

Our empirical analysis uses Social Security Administration records on annual earnings from 1970 to 1979 for a sample of trainees who entered the CETA program in 1976, and a comparison group of nontrainees drawn from the March 1976 Current Population Survey. As is usual in any nonexperiment evaluation, the analysis of training effects on employment probabilities is made more difficult

---

[4] The idea of separating out the effects of training on the probability of employment and the level of earnings, conditional on employment, was suggested to us by Lars Muus. An earlier analysis of training effects on employment probabilities is presented by Kaitz (1979).

[5] This conclusion emerges from experimentally determined training effects for participants in the National Supported Work Demonstration, and was pointed out to us in personal communication from Robert LaLonde.

by the fact that trainees are a nonrandom sample of the population. We use a variety of techniques to control for differences between the trainees and controls. In the first section of the paper, we perform a simple comparison between the pre- and post-training employment probabilities of the trainees and the controls. If the probability of remaining employed from one year to the next (the retention probability) and the probability of moving from unemployment to employment (the accession probability) are the same for a given individual, then a simple comparison of relative changes in employment probabilities among the trainees and controls provides a consistent estimate of the training effect in the context of a linear probability model. With state dependence or with nonlinear probability specifications, this technique will not necessarily eliminate permanent differences between trainees and controls. Methods are presented in the third section of the paper to handle both problems.

Before presenting these methods, however, we consider some estimates of the training effect obtained by conditioning on the entire pre-training employment history. That is, we compare the post-training employment outcomes of trainees and controls with exactly the same pre-training history. Overall estimates are obtained by weighting the results for the individual histories by the trainee sample fractions. This simple technique also highlights many of the difficulties involved in using observational data to evaluate the effects of training.

In the third section we explore a number of relatively parsimonious models for the employment histories of the trainees and controls. These all express the log odds of employment in terms of year, individual, state, and training effects. The discrete nature of the data allows for simple goodness-of-fit tests that indicate the relative success of alternative models. In modelling the effects of training, we allow for separate effects on the accession and retention rates of trainees. We also allow for a one-time impact of training on the probability of employment in the year after training. A distinction between the one-time effect and the permanent effect of training is important because, we argue, there are potentially important biases that affect the former but not the latter.

We initially treat the individual-specific variables as fixed effects. The resulting incidental parameter problem renders conventional maximum likelihood estimates inconsistent. For the simplest model of employment probabilities with no state dependence, this problem can be overcome by maximizing a conditional version of the likelihood, as suggested by Raasch (1960) and Chamberlain (1980). Unfortunately, this technique cannot be extended to the more complex models of state-dependence which we believe are necessary to describe the data. We turn instead to a random effects specification. Rather than specify some simple parametric form for the unobserved heterogeneity, however, we treat the distribution of the individual effects as discrete. Our method is similar to and is motivated by Heckman and Singer's (1984) implementation of the technique of nonparametric maximum likelihood.

The results obtained in the first and second sections suggest that CETA participation had a small but significantly positive impact on the post-training employment probabilities of the trainees. The more sophisticated methodology

employed in Section 3, on the other hand, points to somewhat larger increases in trainee employment rates. Irrespective of the methodology employed, we find that the estimated effects are larger for classroom trainees than for other CETA participants. The results also suggest that movements in and out of employment are reasonably well described by a first-order Markov process, conditional on individual heterogeneity. Judging by their fit to the data, relatively simple models of the distribution of individual effects are quite successful in describing the employment histories of both trainees and controls.

1. DATA DESCRIPTION AND COMPARISONS OF CHANGES IN EMPLOYMENT PROBABILITIES

The CETA programs were federally-funded training and employment pro-grams administered through some 450 city, county, and state agencies across the U.S.[6] The Comprehensive Employment and Training Act provided funding for two distinct types of programs: vocational and retraining programs for unem-ployed or disadvantaged workers (members of households with incomes below the federal poverty standard) administered under Title I of the Act; and countercyclical job creation programs for recent job-losers in high unemployment areas, administered under Titles II and VI of the Act. Participants under Title I were mainly disadvantaged workers, and tended to have relatively lower levels of schooling and labor market experience. A majority of these participants were enrolled in classroom training (short-term vocational courses), on-the-job train-ing, and "work experience" programs (subsidized public sector jobs emphasizing work habits and skills development). Participants under Titles II and VI, on the other hand, tended to have characteristics fairly similar to the overall population of unemployed workers. These participants were mainly enrolled in public sector employment programs, which, for the most part, offered little or no formal training, and provided subsidized employment in the local public sector.

Total CETA enrollment in June 1976 was 806,000. At that time, roughly 20 percent of participants were enrolled in classroom training, 25 percent in work experience programs, and 35 percent in public sector employment programs. The costs of the CETA program to the Federal government in 1976 were $882 per participant in the Title I programs, and $3049 per participant in Title II and VI programs.[7]

Table I contains information on the characteristics and employment histories of our sample of 1976 CETA trainees, as well as a comparison sample of individuals in the March 1976 Current Population Survey (CPS). The trainee data are drawn from the *Continuous Longitudinal Manpower Survey* (CLMS). The comparison sample represents a merged file of CPS records with longitudinal earnings information from the Social Security Administration. A more detailed description of the data sources is provided in the Data Appendix.

---

[6]An overview of the CETA programs in place during 1976 is presented in *Employment and Training Report of the President* (1976, pp. 87–103).

[7]Data on enrollments and costs of the CETA programs in 1976 are summarized in *Employment and Training Report of the President* (1977), Tables F-2 and F-3, pp. 262–263.

TABLE I

DEMOGRAPHIC CHARACTERISTICS AND EMPLOYMENT RATES: TRAINEES AND CONTROLS

	Trainees[a] (1)	Classroom Trainees[b] (2)	Nonclassroom Trainees[c] (3)	Eligible CPS[d] (4)	Control Sample[e] (5)
Demographic Characteristics:					
1. Age in 1976	30.8	29.1	31.3	38.9	32.4
2. Education	11.5	11.1	11.6	12.0	12.8
3. Fraction Married in 1976	.50	.48	.51	.83	.78
Employment Rates:[f]					
4. 1970	.81	.77	.83	.89	.88
5. 1971	.83	.82	.83	.89	.89
6. 1972	.87	.85	.88	.90	.90
7. 1973	.88	.86	.89	.91	.91
8. 1974	.87	.82	.88	.91	.92
9. 1975	.81	.81	.81	.90	.91
10. 1976	.93	.86	.94	.90	.90
11. 1977	.87	.86	.87	.88	.90
12. 1978	.83	.85	.82	.84	.87
13. 1979	.78	.79	.77	.76	.80
Sample Size	2153	481	1672	15245	4207

[a] Male CETA trainees 21 years of age or older who entered and completed training in 1976.
[b] Trainees enrolled in classroom training programs.
[c] Trainees enrolled in on-the-job training and public sector employment programs.
[d] Males 21 years of age and older in the March 1976 Current Population Survey, who were in the labor force during the survey week and reported individual and household incomes in 1975 less than $20,000 and $30,000, respectively.
[e] Stratified random sample of eligible CPS, drawn to have the same distribution of potential experience (age minus education) as the trainees.
[f] Proportion of sample reporting positive Social Security earnings.

In this paper we focus exclusively on the effects of CETA participation on adult male trainees. The sample described in the first column of Table I represents male CETA participants who were 21 years of age or older at enrollment, and who entered and left the program in 1976.[8] The "employment rates" in rows 4–13 of the Table give the fraction of the sample reporting nonzero social security earnings in each year from 1970 to 1979.[9] Unfortunately, we have no way of knowing whether an observation of zero earnings in this data represents a year-long spell of unemployment, withdrawal from the labor force, employment in the untaxed sector of the economy, or missing data.[10] This is a

[8] Not all CETA participants actually completed their assigned training program. From the CLMS data we know only the data of enrollment, the date of program termination, and the kind of program into which the participant was enrolled. Evidence from administrative records on end-of-program placements (*Employment and Training Report of the President* (1977, p. 43) suggests that 5–10 percent of participants moved to unsubsidized employment after only "intake, assessment and/or job referral services from CETA." Another 30–40 percent dropped out of training.

[9] Social Security earnings refer to earnings for which the individual (and his employer) paid Social Security taxes.

[10] According to the Social Security Administration approximately 89.1 percent of wage and salary and self-employed workers in 1970 were covered by Old-Age, Survivors, and Disability Insurance (OASDI), and presumably reported Social Security earnings (*Social Security Bulletin Annual Statistical Supplement*, 1983, p. 61). This percentage was 89.3 in 1975, 89.3 in 1977, and 89.8 in 1979. The major group of untaxed employees work in state and local governments.

major limitation of the Social Security earnings records, which nonetheless represents the only source of times-series earnings information for CETA participants.

The second and third columns of Table I present the characteristics of two distinct groups of trainees: participants in classroom training programs, and other trainees.[11] Classroom trainees were slightly younger than other participants and had slightly worse employment records prior to training. After training, however, the classroom trainees appear to have fared as well as or better than other participants.

For comparison with these trainee samples, the fourth column of Table I presents demographic characteristics and employment histories for a sample of adult males in the March 1976 Current Population Survey. This sample, which we designate the "eligible" CPS sample, includes only those adult males who were in the labor force during the week of the Current Population Survey and who reported individual and household incomes for 1975 of less than $20,000 and $30,000 respectively.[12] Evidently, CETA trainees differ from other members of the population in terms of age, education, marital status, and employment history. Since the trainees are younger, however, a larger fraction of the trainee group may have been out of the labor force in the pre-training period. To control for this important difference in years of labor force attachment, we have drawn a stratified random sample of the CPS with the same distribution of potential labor market experience (age minus education) as the trainees.[13] The characteristics of this sample, which we designate as the "control sample," are displayed in the fifth column of Table I.

From 1970 to 1977 the fraction of the eligible and control samples with nonzero Social Security earnings was approximately constant and equal to the fraction of paid workers in the economy covered by the Social Security system. In 1978 and 1979, however, this fraction fell sharply. A similar decline occurred among both groups of trainees. We believe that this measured decline is due to long delays in filing and recording Social Security earnings.[14] In the empirical analysis reported below we control for reporting delay and other sources of

---

[11] Nonclassroom trainees include participants in on-the-job training programs, work experience programs, and public-sector employment programs. The classroom trainees were mainly (over 95 percent) funded under Title I of CETA. The nonclassroom trainees were mainly (63 percent) funded under Titles II and VI of CETA.

[12] This comparison sample was provided to us by SRI International and was used by Dickinson, Johnson, and West (1984) and Ashenfelter and Card (1985) to analyze the 1976 cohort of CETA trainees. The restrictions on labor force status and individual and household income eliminate approximately 21 percent of the overall CPS sample of adult males.

[13] The trainee and eligible CPS samples were divided into 26 potential experience categories: $0, 1, 2, \ldots, 14$ years; 15–16 years; 17–18 years; 19–20 years; 21–22 years; 23–24 years; 25–27 years; 28–30 years; 31–34 years; 35–38 years; 39–43 years; and 44 or more years. The experience distributions of trainees and eligible CPS members were computed, and then the control sample was drawn from the eligible CPS sample by random sampling within experience strata so as to generate the largest possible control sample with the same distribution of potential experience as the trainee sample.

[14] The earnings information for both trainees and controls was updated in October 1983 and represents the most recent publicly available data.

TABLE II

AVERAGE EMPLOYMENT PROBABILITIES: TRAINEES VERSUS CONTROLS

(Standard Errors in Parentheses)

	Controls	Trainees	Classroom Trainees	Nonclassroom Trainees
1. Average 1970–74	.900	.853	.821	.862
	(.004)	(.006)	(.013)	(.006)
2. Average 1970–75	.901	.846	.819	.854
	(.003)	(.005)	(.012)	(.006)
3. Average 1977–79	.855	.826	.833	.823
	(.004)	(.006)	(.014)	(.007)
4. Difference:	−.046	−.021	.014	−.031
1977–79 less 1970–75	(.004)	(.007)	(.016)	(.008)
5. Difference:	−.045	−.027	.011	−.039
1977–79 less 1970–74	(.004)	(.008)	(.017)	(.009)
6. Difference Relative to Controls:	—	.025	.060	.015
1977–79 less 1970–75		(.008)	(.017)	(.009)
7. Difference Relative to Controls:	—	.018	.056	.006
1977–79 less 1970–74		(.009)	(.017)	(.009)

NOTE: See notes to Table I. Standard errors are based on sample variances of the averages and differences reported in each row.

year-to-year variation in the fraction of nonzero earnings by a series of year effects, which we assume to have identical effects on the employment probabilities of the trainees and the controls.

Some indication of the relative changes in employment probabilities is provided in Table II, which compares pre-1976 and post-1976 probabilities for trainees and controls. For the control group and for the trainees as a whole, average employment probabilities fell after 1976. The drop is smaller for the trainees, providing some evidence of a positive training effect. Comparing Columns 3 and 4 of the table, it is clear that most of the improvement in the trainees' relative position is concentrated among the classroom trainees. Nonclassroom trainees' employment levels follow those of the control sample rather closely.

An obvious question is whether the simple technique presented in Table II of comparing relative changes in employment probabilities between trainees and controls leads to a consistent estimate of the training effect in the presence of unobserved differences between the two groups. Ashenfelter (1978) showed that a similar comparison of relative changes in earnings for trainees and controls leads to a consistent program estimate for the *level* of earnings provided that (1) shocks in pre-training and post-training earnings are uncorrelated with their own lagged values and with the decision to participate in training and (2) the unobserved individual effects enter linearly into the earnings equation.

The equivalent formulation of employment probabilities expresses P_{it}, the probability that individual i is employed in period t, as a simple linear components-of-variance structure:

$$(1) \qquad P_{it} = \alpha_i + \beta_t + \theta D_{it},$$

where α_i is an individual-specific component, β_t is a year-specific component, D_{it} is an indicator for post-1976 trainee status, θ is the training effect, and employment outcomes are assumed to be independent (conditional on the individual effect) across years. Suppose further that pre-training employment outcomes are independent of D_{it}, conditional on the individual effects α_i. Let E_{it} represent an indicator variable whose value is equal to unity if i is employed in t, and let $\bar{E}_{i1}$ and $\bar{E}_{i2}$ represent the individual-specific means of E_{it} in the pre-training (1970–75) and post-training (1977–79) periods, respectively. Then the expected change in average employment rates for the trainees from the pre-training to the post-training period is

$$E\left(\bar{E}_{i2} - \bar{E}_{i1} | D_{i76} = 1\right) = \beta^* + \theta,$$

where β^* represents the difference in the means of the year effects before and after training. The expected change in employment rates for the controls, on the other hand, is

$$E\left(\bar{E}_{i2} - \bar{E}_{i1} | D_{i76} = 0\right) = \beta^*.$$

The expected "difference-in-differences" of the average pre- and post-training employment rates for trainees and controls is

$$E\left(\bar{E}_{i2} - \bar{E}_{i1} | D_{i76} = 1\right) - E\left(\bar{E}_{i2} - \bar{E}_{i1} | D_{i76} = 0\right) = \theta.$$

A consistent estimate of θ is therefore provided by the corresponding difference in sample average changes in pre- and post-training employment rates for trainees and controls.[15] For CETA classroom participants, the difference-in-difference estimate in row 6 of Table II is 6.0 percentage points, with a standard error of 1.7, while the estimate for nonclassroom participants is 1.5 points, with a standard error of 0.9.

There are a number of difficulties with the assumptions leading to the difference-in-difference estimators presented in Table II. First, experience suggests that models that explicitly incorporate the restrictions $P_{it} \geq 0$ and $P_{it} \leq 1$ tend to better summarize discrete data than the linear probability model (1). For example, models such as

$$(2) \qquad F^{-1}(P_{it}) = \alpha_i + \beta_t + \theta D_{it},$$

where F is a logistic or Gaussian distribution function are guaranteed to produce fitted probabilities inside the unit interval, while (1) is not. Unfortunately, the simple differencing schemes employed here will not completely eliminate the individual effects in a nonlinear model such as (2). In the third section of this paper, we develop several models that incorporate a logistic specification for the employment probabilities and that allow for individual effects.

A more fundamental difficulty is the assumption that the participation decision is independent of pre-training outcomes. Various authors have pointed out that

---

[15] The difference-in-differences estimator is not the most efficient linear estimator of θ in (1). Fully efficient estimation of (1) requires a weighted least squares approach.

pre-training earnings may be contaminated by transitory shocks that contributed toward the decision to enter training.[16] Evidence of this shows up in Table I as a sharp decline in trainee employment rates in 1975. This evidence suggests that 1975 employment probabilities are not independent of training status, given the individual effects. If we maintain the structure of equation (1), and assume that employment probabilities prior to 1975 are independent of training status, then a consistent program estimate is obtained by forming the difference-in-differences without 1975 data. Such estimates are shown in row 7 of Table II. For both groups of trainees, the estimated training effects are smaller when 1975 data are excluded. For the classroom group the new estimate is 5.7 points with a standard error of 1.7, while for the nonclassroom trainees the estimate drops to essentially zero.

This procedure can also be repeated to test the independence of training status and pre-1975 employment probabilities.[17] For example, if 1974 data are omitted from the pre-training averages, the estimated training effects change only slightly to .058 and .004, respectively, for classroom and nonclassroom trainees. These results suggest that 1974 employment probabilities are not significantly biased by transitory effects that contributed toward the decision to enter training in 1976.

A final difficulty with the specification of equation (1) is the assumption that employment probabilities are independent of previous employment outcomes. A simple tabulation of average retention probabilities (the probability of employment conditional on employment last period) and average accession probabilities (the probability of employment conditional on unemployment last period) shows that these two are very different. Moreover, evidence in the third section of this paper suggests that accession probabilities are lower than retention probabilities for the same individual. This suggests a lack of independence in employment probabilities over time that complicates the interpretation of the simple difference-in-differences estimators. To begin with, if there is individual-specific heterogeneity (i.e., if the α_i vary across people), then the expectation of the average employment probability for a fixed sample of individuals in a particular year depends on the distribution of individuals between employment and unemployment in the previous year. This phenomenon is especially relevant to the transitions immediately after training. During the training period many more trainees than usual may be counted as employed, thereby increasing the expected employment probabilities immediately after training for CETA participants. On the other hand, many jobs held by CETA trainees in 1976 were automatically terminated with the end of program participation. Thus, state dependence in employment probabilities together with the unknown effects of program participation on employment status in late 1976 introduce unknown biases in the post-training employment probabilities of the trainees.

[16] The phenomenon of a relative dip in pretraining earnings for participants in subsidized training programs was first pointed out by Ashenfelter (1975) and has been confirmed by subsequent analysts, including Kiefer (1979), Bassi (1983), and LaLonde (1986).

[17] A similar specification test was suggested by Ashenfelter (1978) in connection with linear components-of-variance models of the level of earnings.

State dependence also implies that selection bias in the program estimates cannot be eliminated by simply dropping 1975 data, even if 1975 employment status is the only determinant of program participation. Given a particular value for the individual effect, and given that the probability of training is higher for those unemployed in 1975, the probability of employment is lower for trainees in *all* previous years, since unemployed workers in 1975 are more likely to have experienced unemployment in the previous years. In Section 3, we present estimators that allow for selection bias with state dependence as well as individual-specific heterogeneity.

2. COMPARISONS OF EXACT MATCHES

As motivation for the estimates presented in this section, observe that if assignment to training were random, then unbiased and consistent estimates of the effect of training could be obtained by simply comparing trainee and control post-training employment rates. Such estimates obtained from our data would seem most unreliable: CETA participants are obviously a nonrandom sample of the population and we therefore suspect that they differ from the controls in ways other than participation in training. Indeed, this suspicion can be immediately confirmed by an examination of the pre-training data—trainees had considerably worse employment histories prior to training than did controls.

While it clearly does not make sense to directly compare the post-training employment rates of trainees and controls who had markedly different pre-training employment histories, comparisons of trainees and controls with exactly the same pre-training history have a definite intuitive appeal. Such comparisons are made in Table III. Each row of the table corresponds to a different pre-training history. For each history we have calculated the average post-training employment rates for the controls and the trainee groups. The estimated training effect for a given history is simply the difference of the trainee and control employment rates. The total effect (in the last row of the Table) is calculated by weighing the results for the individual histories by the trainee sample fractions. There are four histories, containing a total of 12 trainees, with no observations from the control group. For purposes of calculating average training effects we have ignored these cells and reweighted the remaining cells accordingly.

Compared to the difference-in-difference estimates, the estimates in Table III show a smaller effect for classroom participants (.036 with a standard error of .013 versus .056 with a standard error of .017) and a larger effect for nonclassroom participants (.012 with a standard error of .009 versus .006 with a standard error of .009), although the overall training effects from the two methods are identical. Both methods attribute the larger training effect to classroom participants.

The overall estimates do not, however, convey all the information in Table III. Turning to the individual pre-training histories, it is evident that there is a great deal of variation in the size and even the direction of the training effect. For instance, for trainees with a 000000 employment history (corresponding to no

TABLE III

AVERAGE POST-TRAINING EMPLOYMENT RATE[a] AND TRAINING EFFECT[b]
BY PRE-TRAINING EMPLOYMENT HISTORY

(Standard Errors[c] in Parentheses)

Job History	Controls		Classroom			Nonclassroom			All Trainees		
	N	Rate[a]	N	Rate	Effect[b]	N	Rate	Effect	N	Rate	Effect
000000	109	.099	13	.769	.671	32	.635	.538	45	.674	.576
		(.024)		(.084)	(.088)		(.076)	(.079)		(.059)	(.064)
000001	27	.716	19	.895	.179	23	.754	.038	42	.817	.101
		(.073)		(.050)	(.089)		(.074)	(.105)		(.048)	(.088)
000010	3	.444	2	.667	.222	4	.917	.472	6	.833	.389
		(.240)		(.236)	(.336)		(.072)	(.251)		(.104)	(.262)
000011	21	.841	2	.833	−.008	12	.806	−.036	14	.810	−.032
		(.053)		(.118)	(.129)		(.073)	(.090)		(.065)	(.084)
000100	7	.190	2	.000	−.190	3	.333	.143	5	.200	.010
		(.092)		(.000)	(.092)		(.272)	(.287)		(.179)	(.201)
000101	4	.667	3	.556	−.111	0	—	—	3	.556	−.111
		(.118)		(.240)	(.267)					(.240)	(.267)
000110	1	.667	2	.667	.000	7	.571	−.095	9	.593	−.074
		(.000)		(.236)	(.236)		(.111)	(.111)		(.102)	(.102)
000111	37	.865	5	1.000	.135	24	.819	−.045	29	.851	−.014
		(.035)		(.000)	(.035)		(.055)	(.0650)		(.048)	(.059)
001000	4	.333	1	.667	.333	2	.000	−.333	3	.222	−.111
		(.204)		(.000)	(.204)		(.000)	(.204)		(.181)	(.273)
001001	4	1.000	1	.000	−1.000	1	.667	−.333	2	.333	−.667
		(.000)		(.000)	(.000)		(.000)	(.000)		(.236)	(.236)
001010	0	—	0	—	—	1	.000	—	1	.000	—
							(.000)			(.000)	
001011	3	.889	0	—	—	3	.889	.000	3	.889	.000
		(.091)					(.091)	(.128)		(.091)	(.128)
001100	1	.667	1	1.000	.333	4	.917	.250	5	.933	.267
		(.000)		(.000)	(.000)		(.072)	(.072)		(.060)	(.060)
001101	1	1.000	1	.667	−.333	4	.583	−.417	5	.600	−.400
		(.000)		(.000)	(.000)		(.182)	(.182)		(.146)	(.146)
001110	7	.571	5	.800	.229	9	.630	.058	14	.690	.119
		(.146)		(.179)	(.231)		(.122)	(.191)		(.104)	(.179)
001111	71	.958	10	.833	−.124	39	.829	−.129	49	.830	−.128
		(.013)		(.108)	(.109)		(.047)	(.048)		(.043)	(.045)
010000	3	.667	0	—	—	0	—	—	0	—	—
		(.157)									
010001	1	1.000	0	—	—	1	.333	−.667	1	.333	−.667
		(.000)					(.000)	(.000)		(.000)	(.000)
010010	1	.333	0	—	—	0	—	—	0	—	—
		(.000)									
010011	3	.556	1	.667	.111	2	1.000	.444	3	.889	.333
		(.240)		(.000)	(.240)		(.000)	(.240)		(.091)	(.257)
010100	0	—	1	.333	—	1	.333	—	2	.333	—
				(.000)			(.000)			(.000)	
010101	2	.000	0	—	—	0	—	—	0	—	—
		(.000)									
010110	1	1.000	0	—	—	1	1.000	.000	1	1.000	.000
		(.000)					(.000)	(.000)		(.000)	(.000)
010111	11	.758	1	1.000	.242	3	.889	.131	4	.917	.159
		(.114)		(.000)	(.114)		(.091)	(.146)		(.072)	(.135)
011000	3	.333	1	1.000	.667	3	.778	.444	4	.833	.500
		(.272)		(.000)	(.272)		(.091)	(.287)		(.083)	(.285)
011001	4	.583	1	1.000	.417	6	.889	.306	7	.905	.321
		(.182)		(.000)	(.182)		(.064)	(.193)		(.057)	(.190)

TABLE III (*Continued*)

Job History	Controls		Classroom			Nonclassroom			All Trainees		
	N	Rate[a]	N	Rate	Effect[b]	N	Rate	Effect	N	Rate	Effect
011010	2	.333	0	—	—	2	.667	.333	2	.667	.333
		(.236)					(.000)	(.236)		(.000)	(.236)
011011	8	.875	1	.333	− .542	1	.333	− .542	2	.333	− .542
		(.082)		(.000)	(.082)		(.000)	(.082)		(.000)	(.082)
011100	0	—	2	.500	—	3	.333	—	5	.400	—
				(.118)			(.157)			(.112)	
011101	1	.000	2	.167	.167	7	.810	.810	9	.667	.667
		(.000)		(.118)	(.118)		(.092)	(.092)		(.117)	(.117)
011110	8	.292	9	.778	.486	13	.615	.324	22	.682	.392
		(.124)		(.117)	(.171)		(.088)	(.152)		(.073)	(.144)
011111	149	.888	26	.872	− .016	76	.860	− .028	102	.863	− .025
		(.019)		(.052)	(.055)		(.031)	(.037)		(.027)	(.033)
100000	32	.146	2	.667	.521	9	.630	.484	11	.636	.491
		(.044)		(.000)	(.044)		(.122)	(.130)		(.100)	(.109)
100001	6	.833	1	.667	− .167	4	.750	− .083	5	.733	− .100
		(.104)		(.000)	(.104)		(.138)	(.173)		(.112)	(.235)
100010	1	.333	1	.333	.000	0	—	—	1	.333	.000
		(.000)		(.000)	(.000)					(.000)	(.000)
100011	5	.800	2	.333	− .467	3	.667	− .133	5	.533	− .267
		(.119)		(.236)	(.264)		(.272)	(.297)		(.202)	(.235)
100100	0	—	0	—	—	4	.333	—	4	.333	—
							(.118)			(.118)	
100101	1	.333	0	—	—	4	.917	.583	4	.917	.583
		(.000)					(.072)	(.072)		(.072)	(.072)
100110	6	.500	1	1.000	.500	3	.333	− .167	4	.500	.000
		(.204)		(.000)	(.204)		(.157)	(.258)		(.186)	(.276)
100111	24	.875	1	.000	− .875	19	.789	− .086	20	.750	− .125
		(.047)		(.000)	(.047)		(.071)	(.085)		(.078)	(.091)
101000	3	.778	0	—	—	1	.333	− .444	1	.333	− .444
		(.091)					(.000)	(.091)		(.000)	(.091)
101001	1	1.000	1	.000	− 1.000	1	.333	− .667	2	.167	− .833
		(.000)		(.000)	(.000)		(.000)	(.000)		(.118)	(.118)
101010	1	1.000	0	—	—	1	.333	− .667	1	.333	− .667
		(.000)					(.000)	(.000)		(.000)	(.000)
101011	1	.667	0	—	—	6	.722	.056	6	.722	.056
		(.000)					(.165)	(.165)		(.165)	(.165)
101100	1	.000	0			4	.750	.750	4	.750	.750
		(.000)					(.138)	(.138)		(.138)	(.138)
101101	1	.667	1	1.000	.333	5	.400	− .267	6	.500	− .167
		(.000)		(.000)	(.000)		(.174)	(.174)		(.171)	(.171)
101110	5	.733	3	.333	− .400	11	.667	− .068	14	.595	− .138
		(.112)		(.272)	(.294)		(.113)	(.159)		(.113)	(.158)
101111	67	.866	11	.970	.104	33	.859	− .007	44	.886	.021
		(.032)		(.029)	(.043)		(.038)	(.050)		(.030)	(.044)
110000	21	.190	2	.000	− .190	9	.667	.475	11	.545	.355
		(.073)		(.000)	(.073)		(.105)	(.128)		(.115)	(.137)
110001	9	.741	1	.667	− .074	3	.778	.037	4	.750	.009
		(.136)		(.000)	(.136)		(.091)	(.164)		(.072)	(.154)
110010	2	.167	1	1.000	.833	2	.167	.000	3	.444	.278
		(.118)		(.000)	(.118)		(.118)	(.168)		(.240)	(.267)
110011	17	.804	2	1.000	.196	8	.833	.029	10	.867	.063
		(.088)		(.000)	(.088)		(.059)	(.106)		(.052)	(.102)
110100	4	.583	1	.000	− .583	1	.333	− .250	2	.167	− .417
		(.072)		(.000)	(.072)		(.000)	(.072)		(.117)	(.138)
110101	5	.800	0	—	—	3	.889	.089	3	.889	.089
		(.073)					(.091)	(.117)		(.091)	(.117)

TABLE III (*Continued*)

Job History	Controls		Classroom			Nonclassroom			All Trainees		
	N	Rate[a]	N	Rate	Effect[b]	N	Rate	Effect	N	Rate	Effect
110110	5	.600	1	1.000	.400	5	.866	.267	6	.889	.289
		(.174)		(.000)	(.174)		(.073)	(.189)		(.064)	(.185)
110111	43	.922	7	.952	.030	18	.778	−.145	25	.827	−.096
		(.033)		(.044)	(.055)		(.079)	(.085)		(.060)	(.068)
111000	23	.406	5	.533	.128	9	.704	.298	14	.643	.237
		(.094)		(.202)	(.223)		(.097)	(.135)		(.098)	(.136)
111001	14	.762	1	.000	−.762	9	.704	−.058	10	.633	−.129
		(.071)		(.000)	(.071)		(.110)	(.131)		(.120)	(.139)
111010	3	.556	2	.333	−.222	7	.810	.254	9	.704	.148
		(.240)		(.236)	(.336)		(.132)	(.274)		(.133)	(.274)
111011	40	.892	6	.889	.003	15	.867	−.025	21	.873	−.019
		(.034)		(.101)	(.107)		(.061)	(.070)		(.053)	(.063)
111100	36	.463	7	.524	.061	24	.694	.231	31	.656	.193
		(.068)		(.148)	(.163)		(.081)	(.106)		(.072)	(.099)
111101	27	.802	14	.690	−.112	23	.725	−.078	37	.712	−.091
		(.068)		(.079)	(.104)		(.064)	(.093)		(.050)	(.084)
111110	106	.538	28	.857	.319	136	.814	.276	164	.821	.283
		(.040)		(.054)	(.068)		(.027)	(.048)		(.024)	(.047)
111111	3200	.918	267	.893	−.026	1005	.864	−.054	1272	.870	−.048
		(.004)		(.015)	(.015)		(.008)	(.009)		(.007)	(.008)
TOTAL[d]	4207		481		.036	1663		.012	2153		.018
					(.013)			(.009)			(.008)

[a] Average employment rate in the post-training period (1977–79).
[b] Difference between trainee and control rates.
[c] Standard errors are the maximum likelihood estimates under the assumption of random sampling from an unrestricted multinomial for the eight possible post-training outcomes.
[d] Total effect is the weighted average of the effects for the individual histories using the trainee sample fractions as weights.

Social Security earnings in the entire pre-training period) the estimated training effects are very large. Recall, however, that zero Social Security earnings can mean that a worker is unemployed, that he is employed in the uncovered sector, or even that his data are missing.[18] Thus one explanation for the size of the training effect for workers with the 000000 history is that a large fraction of the controls with this pre-training history were actually in the uncovered sector. Most of these workers would be expected to remain in the uncovered sector after training, thereby depressing their measured employment rates in the post-training period.

The estimated effects for trainees with a 111111 employment history (corresponding to positive Social Security earnings in every period) are, on the other hand, actually negative. It is quite likely, however, that many trainees in this category suffered some setback before they elected to enter training. Trainees who became unemployed part way through 1975, for example, are still recorded as employed in 1975. In the absence of training such individuals would be expected to have lower post-training employment rates than other individuals

[18] We believe that the extent of missing Social Security earnings data is relatively low in the period before 1976.

TABLE IV

ESTIMATED EMPLOYMENT RATE TRAINING EFFECTS BASED ON EXACT MATCHES

(Standard Errors in Parentheses)

	All Trainees	Classroom	Nonclassroom
1. 1977	.026	.025	.027
	(.009)	(.015)	(.010)
2. 1978	.0004	.027	−0.007
	(.010)	(.016)	(.011)
3. 1979	.026	.056	.018
	(.011)	(.018)	(.012)
4. Average Post-Training	.018	.036	.012
	(.008)	(.013)	(.009)

NOTE: See notes to Table III.

with the 111111 history and the training effect for this history is therefore biased downward.

Similar interpretations can be offered for many of the other histories. In particular, the comments about the 000000 history also seem to apply to histories 100000, 110000, and 111000, while those for the 111111 history may also apply to histories 011111, 001111, and 000111. In these cases and others we have reason to suspect that there are important differences between trainees and controls with exactly the same pre-training histories. We cannot, however, verify these suspicions using the present data. For a given history, trainees and controls are identical with respect to all measured characteristics.

One might hope that the overall training effects would be insensitive to any individual history. Unfortunately, this is not the case. If the 000000 history is deleted from the totals, the training effect declines substantially. For classroom trainees it becomes .018, while for nonclassroom trainees it becomes −.006. On the other hand, without the 111111 history, the overall effects increase to .114 for classroom and .123 for nonclassroom participants. This sensitivity must be kept in mind when interpreting the results of this or any observational study of training effects.

Table IV contains separate exact match estimates of the training effect for each post-training year. To save space only the total effects are shown.[19] As can be seen, there is substantial variation in the effect for the three post-training years. For trainees as a whole the effects are .026 with a standard error of .009 in 1977, .0004 with a standard error of .010 in 1978, and .026 with a standard error of .011 in 1979. The time pattern is quite different, however, among classroom and nonclassroom trainees. While the two groups show approximately equal effects in 1977, the nonclassroom trainees drop to a negative effect in 1978 before recover-

[19] The pattern of training effects across histories for individual years is similar to that of the averages shown in Table III. For instance, the effects for the 000000 history are always large and positive and the effects for the 111111 history are always negative or insignificantly different from zero.

TABLE V

ESTIMATED TRANSITION RATE TRAINING EFFECTS BASED ON EXACT MATCHES

(Standard Errors in Parentheses)

	All Trainees	Classroom	Nonclassroom
1977–78:			
1. Retentions	−.018	−.019	−.024
	(.010)	(.015)	(.011)
2. Accessions	.033	.031	.025
	(.044)	(.067)	(.046)
1978–79:			
3. Retentions	.023	.098	−.015
	(.011)	(.017)	(.012)
4. Accessions	−.014	−.081	−.012
	(.034)	(.062)	(.036)
Pooled 1977–78, 1978–79:			
5. Retentions	.006	.002	−.0005
	(.008)	(.012)	(.008)
6. Accessions	.003	.040	−.009
	(.029)	(.055)	(.031)

NOTE: See text for discussion. Estimates are maximum likelihood for the unrestricted multinomial model for post-training outcomes. Estimated standard errors are obtained by the delta method.

ing somewhat in 1979. The classroom trainees, on the other hand, show a marginal increase in the training effect in 1978 and relatively substantial 5.6 percentage point training effect in 1979.

The technique of exact-match comparisons can easily be extended to measure the separate effects of training on accession and retention rates. By using only post-1977 transitions, we can also control for any one-time effects of training on employment status in the year immediately after training. Exact-match comparisons of transition probabilities for trainees and controls are presented in Table V. The training effect reported for the 1977–78 retention rate, for example, compares the retention rates between 1977 and 1978 for trainees and controls with identical pretraining histories. The differences in retention rates for each pretraining history are then weighted by the fraction of trainees with each history to arrive at an average estimated training effect.

Formally, let $p(110|h_j)$ represent the probability of the sequence of post-training employment indicators "110" (i.e., employed in 1977, employed in 1978, unemployed in 1979) conditional on the jth history of pre-training employment indicators. Then the retention rate between 1977 and 1978 for the jth history is the probability

$$r_j(77,78) = \frac{p(110|h_j) + p(111|h_j)}{p(110|h_j) + p(111|h_j) + p(100|h_j) + p(101|h_j)}.$$

An estimate of $r_j(77, 78)$ can be obtained from estimates of the sample probabilities of the various post-training outcomes. The difference in the weighted averages of these estimated retention rates between trainees and controls (weighted by the sample fraction of trainees with the jth history) represents the estimated training effect for the 1977–78 retention rate.[20] Standard errors are obtained for the average retention and accession rates from the estimated standard errors for the retention and accession rates for each pretraining history. These, in turn, are estimated by the delta method using the sampling variability of the estimated probabilities of the post-training outcomes, conditional on the pretraining history.[21]

An estimate of the *average* post-training retention rate for the jth pretraining history is obtained by forming a weighted average of $r_j(77, 78)$ and $r_j(78, 79)$. The weights are simply the relative fractions of individuals with the jth pretraining history who were at risk of remaining employed between 1977–78, and 1978–79, respectively. The weighted average retention rate is therefore

$$\bar{r}_j = \frac{p(110|h_j) + 2p(111|h_j) + p(011|h_j)}{\begin{bmatrix} 2p(110|h_j) + 2p(111|h_j) + p(100|h_j) \\ + p(101|h_j) + p(011|h_j) + p(010|h_j) \end{bmatrix}}$$

where each of the probabilities is conditioned on the jth pretraining history.

An estimate of the average training effect on post-training retention rates is obtained by forming a weighted average of the differences between average retention rates of trainees and controls with each pretraining history. Standard errors for the average training effect can be obtained from standard errors for the estimates of the average retention rates for each pretraining history, which in turn are constructed by the delta method from the sampling variability of the estimated probabilities of the various post-training outcomes.[22]

For all three groups of trainees the estimated training effects obtained by pooling the 1977–78 and 1978–79 transitions are smaller than their standard errors. The training effects for the individual years are somewhat poorly determined and tend to change signs between 1977–78 and 1978–79. The classroom trainees show a large positive training effect for the 1978–79 retention rate and a large but imprecisely estimated negative training effect for the 1978–79 accession

[20] A difficulty can arise with this estimator if there are no observations in the control group from which to estimate an accession or retention rate that is actually observed in the trainee sample. When this occurred in our samples we ignored the trainee data and reweighted the trainees with available match-groups accordingly.

[21] The estimated standard errors for the accession rates are considerably larger than those for the retention rates. This is due to the smaller sample sizes for measuring these rates—in any given year more than half the population shows positive Social Security earnings. The problem is compounded by the use of the trainee sample fractions as weights for the computation of the overall effect. Those histories which had relatively more potential accessions in the post-training period tended to be the ones that received small weights.

[22] Estimates of the average post-training accession rates and the corresponding training effects are obtained by similar calculations.

rate. Overall, these results suggest that the main effect of training may have worked through the accession rate of the classroom trainees. This finding is confirmed by a naive difference-in-differences of post-training and pre-training transition rates between trainees and controls.[23] Such an analysis shows a sharp decrease in accession rates for the controls between the pre-training and post-training period (from an average of .36 over the 1970–74 period to an average of .25 in the 1977–79 period) with no corresponding drop for the classroom trainees (an average accession rate of .35 in both periods).

The strategy of these exact match procedures is somewhat different from that of the other methods presented in this paper. Rather than specifying a distribution of employment outcomes in terms of unobservable individual effects, the exact match procedure specifies the most general possible model of post-training outcomes in terms of the observable data (pre-training outcomes and training status). In contrast to the training effect estimates derived from equations (1) or (2), which rely on essentially arbitrary functional form assumptions, the exact match procedure relies on the assumption that the participation decision is independent of any unobservable determinants of the probability of employment, conditional on the observable pretraining data.[24] As Heckman and Robb (1986) have observed, this is a strong (and in the context of the model, untestable) assumption. In the next section we present a model of training that allows for a limited form of dependence between the unobservable components of the probability of employment and the unobservable determinants of the decision to participate in training.

Finally, we note two other difficulties with the exact match procedures presented here. First, exact match estimators of the training effect use many degrees of freedom, and result in potentially inefficient estimates. Second, the exact match methodology cannot be used when there are continuous covariates, and would be impractical when there were many more discrete variables than we have here. Indeed, even in the present application, the sample sizes for many of the individual histories are too low to give useful results. Nevertheless, *approximate* match methods which group similar individuals into the same cells might still prove useful in some program evaluation settings.[25]

---

[23] Differences-in-differences of transition rates must be interpreted cautiously since the expected values of the transition rates depend on the distribution of individuals between employment states in the preceding year, and these distributions may in turn depend on training status.

[24] In this regard, the exact-match procedure is analogous to methods of program evaluation for the level of earnings that simply regression-adjust for all observable characteristics (including pre-training earnings). Such methods are described in Goldberger (1972) and their relative performance is considered by LaLonde (1986).

[25] Rosenbaum and Rubin (1985a, 1985b) have recently advocated the use of the propensity score (the conditional probability of assignment to treatment given a vector of observed covariates) in constructing approximate matches. Heckman and Robb (1986) discuss the limitations of this technique and its relation to more familiar methods. LaLonde and Maynard (1987) explore the relative success of matched comparison group estimators of the effect of training for National Supported Work Demonstration data (where a true experimental control group is available). They conclude that the matched comparison group methods are generally no better (and in some cases clearly worse) than other program estimators.

3. NONLINEAR MODELS FOR THE EFFECT OF TRAINING ON EMPLOYMENT PROBABILITIES

In Section I we noted that when employment probabilities are modelled as nonlinear functions of the individual effects, or when state dependence is allowed, the simple difference-in-differences estimator is not necessarily consistent for the training effect. In this section we present estimators of the training effect that allow for these complications. Specifically, we present a logistic regression model of the employment probabilities that includes individual effects and state dependence effects. We also present a model of participation in training that permits interactions between the unobservable components of the employment probabilities and the individual-specific determinants of training status.

We first present a logistic regression model that assumes independence over time in successive employment probabilities. Although this model incorporates a very general specification of the individual effects, it provides a relatively poor fit to either the control group or the trainee data. We then go on to present a class of logistic regression models that include state dependence and a random-effects specification of individual heterogeneity. We find that these models are much more successful in describing the employment histories of the control group. This gives us somewhat more confidence in their application to the problem of determining training effects from nonexperimental data.

As a starting point, consider a model for the controls that assumes independence of successive employment probabilities and is linear in the log-odds of employment:

$$(3) \qquad \text{Logit}(P_{it}) = \alpha_i + \beta_t,$$

where $\text{Logit}(z) = \log(z/(1-z))$ is the inverse logistic distribution function, and P_{it} is, as before, the probability that individual i is employed in period t. For a sample of T observations on each of N individuals, this model can be estimated by maximum likelihood, treating α_i and β_t as parameters. It can be shown, however, that the resulting estimates are inconsistent as the number of individuals (N) tends to infinity. The problem is that the number of parameters $(N + T - 1)$ tends to infinity with the size of the sample.[26]

Raasch (1960), Andersen (1973), and Chamberlain (1980) show that consistent estimates of β_t can be obtained by maximizing a conditional version of the likelihood function in which the likelihood of a given employment sequence is calculated conditional on the total number of years of positive earnings in the sequence. In particular, equation (3) implies that the likelihood of a sequence of employment indicators $\{E_{i1}, E_{i2}, \ldots, E_{iT}\}$ for individual i, conditional on $S_i = \Sigma_t E_{it}$ is:

$$(4) \qquad f(E_{i1}, E_{i2}, \ldots, E_{iT}|S_i) = \frac{\exp\left(\Sigma_t E_{it}\beta_t\right)}{\Sigma_{d \in D(S_i)} \exp\left(\Sigma_t d_t \beta_t\right)}$$

---

[26] See Chamberlain (1980, p. 228) for a more complete discussion.

where $D(S_i)$ is the set of alternative sequences of employment indicators with exactly S_i years of positive earnings.[27] Since the number of "successes" is, for every fixed set of β's, a sufficient statistic for α_i in the logistic regression model (3), the conditional likelihood does not depend on α_i. Consistent estimates of the year effects may therefore be obtained by maximizing equation (4).

In the absence of individual-specific time-varying covariates in equation (3), the right-hand side of equation (4) is constant for every individual with a given sequence of employment indicators. Maximization of the conditional log-likelihood is therefore equivalent to maximizing

$$\sum_{s=0}^{T} \sum_{k \in D(s)} n_{k,s} \log \Pi_{k,s}(\beta),$$

where $n_{k,s}$ denotes the number of individuals with the kth employment history in the sth sufficiency class (i.e., with the same number of years of nonzero earnings) and $\Pi_{k,s}(\beta)$ is the predicted probability of the kth alternative within the sth sufficiency class, as determined by (4).

An appropriate goodness-of-fit statistic for the model of equation (3) is therefore the likelihood ratio statistic

$$2 \sum_{s=0}^{T} \sum_{k \in D(s)} n_{k,s} \log\left(p_{k,s} / \Pi_{k,s}(\hat{\beta}) \right),$$

where $p_{k,s}$ represents the fraction of observations with the kth employment history in the sth sufficiency class, and $\hat{\beta}$ is the vector of conditional maximum likelihood estimates of the year effects.[28] For the case of the 10 year employment histories of the control sample, the degrees of freedom of the test statistic are 1024 minus the number of sufficiency classes (11) minus the number of estimated year effects (9).[29]

The value of this test statistic for the fit of equation (4) to the control group data is reported in the first column of Table VI. Evidently, the model does a relatively poor job of describing the distribution of employment outcomes among the controls. Inspection of the model's residuals suggests that a major difficulty is the inability to predict serial correlation in the observed sequences of zeros and ones. Except for the influence of the year effects, the model predicts that sequences of employment indicators with the same total number of years of employment are equally likely. As the data in Table III show, however, a serially correlated sequence of indicators like 000111 or 111000 is far more likely than an

---

[27] There are $\binom{T}{S_i}$ elements in $D(S_i)$. Note that the sequences $(0,0,\ldots,0)$ and $(1,1,\ldots,1)$ are the only ways of getting zero and 10 successes, respectively. These sequences therefore have conditional likelihoods of unity.

[28] See Kendall and Stuart (1973, pp. 436–437).

[29] We are simultaneously fitting multinomial distributions to the sets of alternative sequences within each of the sufficiency classes. The degrees of freedom within each sufficiency class is the number of alternatives in that class, minus one. Adding over the 11 sufficiency classes, the degrees of freedom is 1024 minus 11 minus the number of estimated parameters.

TABLE VI

GOODNESS-OF-FIT, ESTIMATED TRAINING EFFECTS AND IMPLIED CHANGES
IN THE PROBABILITY OF EMPLOYMENT: CONDITIONAL LOGIT MODEL[a]

	Controls Alone	Controls and All Trainees	Controls and Classroom Trainees	Controls and Nonclassroom Trainees
1. Goodness-of-fit				
Control Contribution	2687.6	2697.8	2690.4	2694.5
Trainees Contribution	—	1426.4	516.6	1054.9
Total	2687.6	4124.2	3206.9	3749.3
(Degrees of freedom)	(1004)	(1250)	(1250)	(1250)
2. Estimated Training				
Effect (θ)	—	.467	.860	.350
(Standard error)		(.097)	(.164)	(.104)
3. Estimated Increase in Employment Probabilities Due to Training[b]				
(a) 1977	—	.037	.069	.027
(b) 1978	—	.043	.078	.032
(c) 1979	—	.062	.114	.046
(d) Average 1977–79	—	.047	.087	.035

[a] Estimated on 1970–79 employment outcomes for the control sample and 1970–74 and 1977–79 outcomes for the trainee samples.
[b] Weighted average of predicted changes in employment probabilities within sufficiency classes.

alternating sequence like 001011 or 101010.[30] An obvious explanation for this finding is that individual retention probabilities are significantly higher than accession probabilities. Individuals who are employed or unemployed are therefore more likely to remain in their previous state in the next year.

Before turning to models that incorporate state dependence, however, we give the results of extending the model of equation (3) to the trainees. The complete model can be written as

$$(5) \qquad \text{Logit}(P_{it}) = \alpha_i + \beta_t + \theta D_{it},$$

where D_{it} is an indicator for post-1976 trainee status and the other variables are the same as in (3). Following the discussion in Section I, we also assume that the probability that an individual entered training depends on his 1975 employment status and/or the value of his fixed effect. In the absence of state dependence, this assumption implies that pre-1975 and post-1976 employment outcomes are independent of the decision to enter training, conditional on α_i. The model can then be estimated by maximizing the conditional likelihood function, using employment outcomes from 1970–74 and 1977–79 for the trainees and from 1970–79 for the controls. The likelihood for the trainees is conditional on the

[30] For example, of the 112 control observations with 3 periods of employment in the first 6 years of the data, 23 have the history 111000, 7 have the history 001110, and 37 have the history 000111. The other 45 observations are distributed over the 17 remaining histories with 3 successes in 6 periods.

number of periods of nonzero earnings in 8 years, while the likelihood for the controls is conditional on the number of periods of nonzero earnings in 10 years.

Estimation results for equation (5) are presented in columns 2–4 of Table VI. For both groups of trainees the estimated training effect is positive and statistically different from zero. The implied increases in the average post-training employment probabilities are 8.7 and 3.5 percent for the classroom and nonclassroom trainees, respectively. These estimates are somewhat higher than the estimates from either the exact match or difference-in-differences procedures of the previous sections. The increasing magnitude of the estimated training effects over the post-training period is due to the sharp decrease in the estimated year effects in 1978 and 1979. Assuming that the training effect on the log-odds of employment is constant, the effect on the probability of employment is higher, the lower the average probability of employment (provided that the average probability is greater than one-half). The goodness-of-fit statistics for the joint model of the trainees and controls, however, are very unfavorable, suggesting that the estimated training effects must be interpreted cautiously.

State dependence can be introduced into the employment probability model for the controls by including a term in the lagged employment indicator E_{it-1}:

$$(6) \qquad \text{Logit}\,(P_{it}) = \alpha_i + \beta_t + \gamma E_{it-1} \quad \text{for} \quad t > 1970.$$

The parameter γ represents the increase in the log-odds of employment in t, conditional on employment in $t-1$. If retention probabilities are higher than accession probabilities then we expect $\gamma > 0$. The model of equation (6) is completed by specifying the distribution of employment probabilities in 1970. For simplicity we assume that the probability of employment in 1970, conditional on α_i, is equal to the "steady-state" employment probability implied by equation (6):[31]

$$(7) \qquad P_{it} = \frac{P_{it}^A}{1 - P_{it}^R + P_{it}^A}, \qquad t = 1970,$$

where P_{it}^A and P_{it}^R refer to the accession and retention probabilities for i in period t, respectively, as determined by equation (6).

Unfortunately, estimation of the model implied by equations (6) and (7) is not as straightforward as estimation of the model implied by equation (3). In the Appendix we show that in the presence of state dependence the minimal sufficient statistic for α_i is, for all but a few exceptional values of the other parameters, the entire data vector for individual i. Thus the conditional likelihood approach cannot be extended to the logistic probability model with state dependence.[32]

---

[31] The term "steady state" is perhaps misleading because we allow an unrestricted year effect in the 1970 employment probability specification.

[32] The situation is analogous to estimation of a linear model of the form $y_{it} = \alpha_i + x_{it}\beta + \gamma y_{it-1} + \varepsilon_{it}$. If $\gamma = 0$, then $\Sigma_t y_{it}$ is a sufficient statistic for α_i and the conditional likelihood approach leads to the usual analysis of covariance (see Chamberlain (1980)). If $\gamma \neq 0$, however, other methods are required to obtain consistent estimates of β.

We turn instead to a random effects specification. That is, we make the additional assumption in equation (6) (and all subsequent models) that the α_i are independent and identically distributed random variables with some common distribution function F. Rather than specify a parametric form for F, however, we assume that F is a discrete distribution with a small number of mass points. We allow the positions of the mass points and the associated probabilities to be parameters of the likelihood function. This specification is intended to be an approximation to the nonparametric maximum likelihood estimator of Kiefer and Wolfowitz (1956). Actual nonparametric maximum likelihood estimates would be obtained by jointly choosing F (unconstrained by any parametric restrictions) and the structural parameters to maximize the likelihood function. Laird (1978) and Lindsay (1983a, 1983b) show that for problems of the type considered here, the maximum will occur at a distribution with finite support. The technique has been applied to econometric models for duration data by Heckman and Singer (1984).

After some experimentation we chose to use four mass points in the distribution function of the individual effects. For example, the first row of Table VII summarizes the results of applying equation (6) to the control data with $\gamma = 0$ and four mass points. The estimated year-effects (not shown in the Table) are identical, to two decimal places, to the estimates obtained from the conditional maximum likelihood procedure summarized in the first column of Table VI. The addition of extra mass points to this model brought only slight increases in the maximized likelihood function, and negligible changes in the estimated year effects.[33] The four estimated mass points, and the estimated fractions of the control group associated with each mass point, are described in the right hand columns of Table VII. The restriction $\gamma = 0$ implies that the estimated log-odds of employment are independent of previous employment status: therefore the log-odds in rows (1a) and (1b) are identical.

The second row of Table VII summarizes the estimates of equation (6) obtained by our random effects technique with γ unrestricted. The addition of one extra parameter for state dependence reduces the goodness-of-fit statistic shown in the second column of Table VII by 1746.3. The new value is actually below the mean of the appropriate chi-squared distribution under the null hypothesis of a correct model. The estimate of γ is 2.75 with an estimated standard error of 0.07.[34] Clearly the model of equation (6) provides a better description of the control group data than the model of equation (3).

The third and fourth rows of Table VII summarize the estimation results for two additional models, both of which allow for interactions between the individual effects and the state effects. The estimates in the third row allow a "one

---

[33] With six mass points the estimated year effects are the same as the conditional likelihood estimates to four decimal places.

[34] The estimation of appropriate standard errors for nonparametric maximum likelihood parameter estimates is unsettled. We report estimated standard errors based on the inverse of the sample information matrix, which are approximate under the assumption that the distribution of the fixed effects is in fact a four-point distribution on a closed and bounded interval.

TABLE VII

SUMMARY OF ALTERNATIVE RANDOM-EFFECTS LOGISTIC MODELS (CONTROL GROUP ONLY)

Model[a]	Goodness-of-Fit Statistic[b] (Degrees of Freedom)	Description of Estimated Mass Points[c]		Type 1	Type 2	Type 3	Type 4
1. No State Dependence $\text{Logit}(P_{it}) = \alpha_i + \beta_t$	2695.4 (1007)	(a)	Log-odds of employment (unemployed last year)	−3.39	−0.34	1.51	4.12
		(b)	Log-odds of employment (employed last year)	−3.39	−0.34	1.51	4.12
		(c)	Estimated weight	0.032	0.065	0.213	0.691
2. One-Parameter State Dependence $\text{Logit}(P_{it}) = \alpha_i + \beta_t + \gamma E_{it-1}$	949.1 (1006)	(a)	Log-odds of employment (unemployed last year)	−5.07	−1.66	−0.48	1.21
		(b)	Log-odds of employment (employed last year)	−2.32	1.09	2.27	3.96
		(c)	Estimated weight	0.017	0.051	0.274	0.659
3. Two-Parameter State Dependence $\text{Logit}(P_{it}) = \alpha_i + \beta_t + (\gamma + \delta\alpha_i)E_{it-1}$	900.4 (1005)	(a)	Log-odds of employment (unemployed last year)	−2.50	−0.53	5.25	11.23
		(b)	Log-odds of employment (employed last year)	2.26	2.30	2.42	2.54
		(c)	Estimated weight	0.080	0.310	0.064	0.545
4. Four-Parameter State Dependence $\text{Logit}(P_{it}) = \alpha_i + \beta_t + \gamma_t \bar{E}_{it-1}$	832.5 (1003)	(a)	Log-odds of employment (unemployed last year)	−2.61	−0.90	22.97	−0.41
		(b)	Log-odds of employment (employed last year)	4.75	0.66	2.81	3.08
		(c)	Estimated weight	0.360	0.048	0.179	0.413

[a] See text for more complete discussion.
[b] Likelihood ratio test statistic against an unrestricted multinomial model of the distribution of control group earnings histories into 1024 cells.
[c] Log-odds are evaluated at the value of the year effect corresponding to 1970.

degree of freedom" interaction:

(8) $\text{Logit}(P_{it}) = \alpha_i + \beta_t + \gamma E_{it-1} + \delta \alpha_i E_{it-1}$ for $t > 1970$,

while the estimates in the fourth row allow a full interaction between the two:

(9) $\text{Logit}(P_{it}) = \beta_t + \alpha_i + \gamma_i E_{it-1}$ for $t > 1970$.

Both models again assume that 1970 employment probabilities for individual i are given by the steady state probabilities corresponding to the transition probabilities for that year.

The log-odds of the retention and accession probabilities for each of the four types are presented in the right-hand columns of Table VII. The model of equation (9) (in row 4 of the Table) imposes no restrictions on the relative transition probabilities, while the other models impose various degrees of constraint. Comparison of the goodness-of-fit statistics suggests that the constrained models do not do a particularly good job of matching the unconstrained fit. The model of equations (6) and (8) can both be easily rejected in favor of the unrestricted model in equation (9).

These goodness-of-fit comparisons suggest that the unrestricted model of equation (9) should be used as the basis for a joint model of the trainees and controls. On the other hand, the goodness-of-fit of the simplest state-dependence model (equation (6)) is acceptable by conventional standards, and the computational burden is considerably lower. We have therefore chosen to use this specification as our basic model.

The extension of the employment probability model represented by equation (6) to the trainee data requires three steps. The first is a specification of the training effects on the employment probabilities. We assume that the effects of training are captured by four parameters: two parameters representing the once-for-all effects of training on employment status in 1977; and two parameters representing the permanent effects of training on the accession and retention rates of the trainees after 1977.

Formally, we assume that the employment probabilities of the trainees are given by

(10a) $\text{Logit}(P_{it}) = \alpha_i + \beta_t + \gamma E_{it-1}$, $1971 \leqslant t \leqslant 1975$,

(10b) $\text{Logit}(P_{i77}) = E_{i75}\{\text{Logit}(P_{i77}|E_{i75} = 1) + \theta_1\}$

$\qquad\qquad\qquad + (1 - E_{i75})\{\text{Logit}(P_{i77}|E_{i75} = 0) + \theta_0\}$,

(10c) $\text{Logit}(P_{it}) = \tau_0 + \alpha_i + \beta_t + (\gamma + \tau_1 - \tau_0)E_{it-1}$, $t \geqslant 1978$.

In equation (10b), $P_{i77}|E_{i75} = 1$ refers to the probability that individual i is employed in 1977, given that he was employed in 1975, while $P_{i77}|E_{i75} = 0$ refers to the probability that i is employed in 1977, given that he was unemployed in 1975.[35] The parameters θ_0 and θ_1 measure the once-for-all effects of training on

---

[35] The probability $P_{i77}|E_{i75} = 1$ is equal to $P_{i76}^R P_{i77}^R + (1 - P_{i76}^R)(P_{i77}^A)$, where P_{it}^A and P_{it}^R are the accession and retention probabilities of individual i in period t, respectively, as determined by equation (10a). The probability $P_{i77}|E_{i75} = 0$ is equal to $P_{i76}^A P_{i77}^R + (1 - P_{i76}^A)P_{i77}^A$.

the log-odds of employment in 1977, conditional on unemployment and employ-ment in 1975 (using equation (10a)), respectively. The parameters τ_0 and τ_1, on the other hand, measure the permanent effects of training on the post-1977 accession and retention rates. We continue to assume that 1970 employment probabilities are equal to the steady state employment probabilities implied by equation (10a).

The second component of the model for the trainee data consists of a model of the determinants of training status in 1976. We assume that the decision to enter training is determined entirely by employment status in 1975 (and perhaps by the value of the unobservable individual effects). In the presence of state dependence, however, it is not sufficient to simply drop the 1975 employment outcome in order to avoid selection bias in the pretraining data. Instead, we explicitly parameterize the dependence of the training decision on employment status in 1975. As a first alternative, we assume that

$$(11) \qquad P(\text{Training}|E_{i75} = 0, \alpha_i) = \rho P(\text{Training}|E_{i75} = 1, \alpha_i),$$

where ρ represents the relative likelihood of entering training from unemploy-ment, as compared to employment. Using the facts that

$$P(E_{i70}, \dots, E_{i75}|\text{Training}, \alpha_i)$$
$$= \frac{P(\text{Training}|E_{i70}, \dots E_{i75}, \alpha_i) \cdot P(E_{i70}, \dots E_{i75}|\alpha_i)}{P(\text{Training}|\alpha_i)}$$

and

$$P(\text{Training}|\alpha_i) = P(\text{Training}|E_{i75} = 1, \alpha_i) \cdot P(E_{i75} = 1|\alpha_i)$$
$$+ P(\text{Training}|E_{i75} = 0, \alpha_i) \cdot P(E_{i75} = 0|\alpha_i),$$

we can write the probability of an observed sequence of pre-training employment outcomes, conditional on training (and the value of the individual effect) as:

$$(12) \qquad P(E_{i70}, \dots E_{i75}|\text{Training}, \alpha_i)$$
$$= \frac{P(E_{i70}, \dots E_{i75}|\alpha_i)}{P(E_{i75} = 1|\alpha_i) + \rho P(E_{i75} = 1|\alpha_i)}, \quad E_{i75} = 1,$$
$$= \frac{\rho P(E_{i70}, \dots E_{i75}|\alpha_i)}{P(E_{i75} = 1|\alpha_i) + \rho P(E_{i75} = 0|\alpha_i)}, \quad E_{i75} = 0.$$

The probabilities in the numerator and denominator of equation (12) may be readily calculated from equation (10a).

This parameterization of the participation decision assumes that selection into training is independent of individual characteristics, conditional on observable employment status in 1975. A more general model is one that allows for differing relative selection probabilities for different values of the individual effect:

$$(13) \qquad P(\text{Training}|E_{i75} = 0, \alpha_i) = \rho(\alpha_i) \cdot P(\text{Training}|E_{i75} = 1, \alpha_i).$$

This model implies that the probability of an observed sequence of pretraining

employment outcomes, conditional on training and the individual effect, is

$$(14) \quad P(E_{i70}, \ldots E_{i75} | \text{Training}, \alpha_i)$$

$$= \frac{P(E_{i70}, \ldots E_{i75} | \alpha_i)}{P(E_{i75} = 1 | \alpha_i) + \rho(\alpha_i) P(E_{i75} = 0 | \alpha_i)}, \quad E_{i75} = 1,$$

$$= \frac{\rho(\alpha_i) P(E_{i70}, \ldots E_{i75} | \alpha_i)}{P(E_{i75} = 1 | \alpha_i) + \rho(\alpha_i) P(E_{i75} = 0 | \alpha_i)}, \quad E_{i75} = 0.$$

In our second model of training status we assume that the $\rho(\alpha_i)$ are completely unrestricted. If the individual effects take on 4 discrete values, this specification introduces 3 extra parameters relative to the selection model of equation (11).

The final step in building a model for the trainee data is the specification of the distribution of the α_i's. One alternative is to allow completely separate discrete distributions for the trainees and controls. On the other hand, a very parsimonious alternative is to model the trainees as a sample from the control group population, with the relative fractions of trainees and controls of each type determined by the trainee selection process. We compromise here and force the mass points to have the same positions for the trainees and controls, while allowing arbitrary sets of weights for the two groups.

The two alternative models of the trainee data were estimated jointly with the model for the control group data (equation (6)) by maximum likelihood. The results are summarized in Tables VIII, IX, and X. For simplicity, we refer to the model for the trainees consisting of equations (10a), (10b), (10c), and (12) as model T1, and to the model consisting of equations (10a), (10b), (10c), and (14) as model T2. These two models differ only in their specification of the relative selection probabilities of employed and unemployed workers into training.

Estimates of the training effects, the state-dependence parameters (γ), and the goodness-of-fit statistics for the two models are presented in Table VIII. We have estimated each model three times: once with the controls and all the trainees; once with the controls and the classroom trainees only; and once with the controls and the nonclassroom trainees. An obvious extension of our analysis would be to model both groups of trainees simultaneously with some model of the relative probability of entering classroom and nonclassroom training.

The estimates of the state dependence parameter γ are very similar across models and across trainee groups and are also very similar to the estimate obtained on the control group alone. The estimated training effects, by comparison, vary somewhat between the models and between the different trainee groups. In general the estimated training effects on the post-1977 accession and retention rates (τ_0 and τ_1) are significantly positive, with the larger effect being on the accession rate. On the other hand, the estimated training effects on employment status in 1977 (θ_0 and θ_1) are significantly positive only for workers who had no Social Security earnings in 1975. For workers with positive earnings in 1975, the estimated training effects on employment status in 1977 are negative and small relative to their standard errors.

TABLE VIII

ESTIMATED TRAINING EFFECTS FOR MODELS T1 AND T2

(Standard Errors in Parentheses)

	All Trainees	Classroom Trainees	Nonclassroom Trainees
Model T1:			
1. Training Effect on 1977 Employment, Conditional on Employment in 1975 (θ_1)	−.074 (.123)	−.067 (.216)	−.085 (.126)
2. Training Effect on 1977 Employment, Conditional on Unemployment in 1975 (θ_0)	.671 (.153)	.999 (.320)	.616 (.168)
3. Training Effect on Post-1977 Retention Rates (τ_1)	.274 (.084)	.591 (.152)	.169 (.088)
4. Training Effect on Post-1977 Accession Rates (τ_0)	.599 (.113)	.907 (.212)	.552 (.124)
5. Estimated State Dependence Parameter (γ)	2.56	2.67	2.64
6. Goodness-of-Fit Statistic[a] (degrees of freedom)			
Controls	968.0 (1023)	954.1 (1023)	965.0 (1023)
Trainees	656.7 (511)	379.7 (511)	528.8 (511)
Total	1624.6 (1509)	1333.8 (1509)	1493.8 (1509)
Model T2:			
7. Training Effect on 1977 Employment, Conditional on Employment in 1975 (θ_1)	−0.34 (.092)	−.039 (.191)	−.024 (.083)
8. Training Effect on 1977 Employment, Conditional on Unemployment in 1975 (θ_0)	.543 (.143)	.922 (.276)	.463 (.155)
9. Training Effect on Post-1977 Retention Rates (τ_1)	.235 (.109)	.578 (.149)	.124 (.084)
10. Training Effect on Post-1977 Accession Rates (τ_0)	.681 (.109)	.936 (.210)	.676 (.115)
11. Estimated State Dependence Parameter (γ)	2.57	2.67	2.61
12. Goodness-of-Fit Statistic[a] (degrees of freedom)			
Controls	967.5 (1023)	954.9 (1023)	964.3 (1023)
Trainees	636.2 (511)	376.5 (511)	509.3 (511)
Total	1603.8 (1506)	1331.4 (1506)	1473.6 (1506)

[a] Degrees of freedom for controls and trainees are unadjusted for parameter estimation. Degrees of freedom for total are adjusted for parameter estimation.

This last result may reflect the potential bias identified in Section 2 with respect to the 111111 pre-training history. That is, many of the individuals who showed positive Social Security earnings in 1975 but who entered training in 1976 may have become unemployed during late 1975 or early 1976. Conditional on employment in 1975, therefore, trainees would have been expected to fare worse than nontrainees in 1977. One of the advantages of the training model presented here is that this kind of bias does not affect the estimated permanent effects of training. It merely complicates the interpretation of the parameters θ_1 and θ_0. In particular, the parameter θ_1 measures not only the effects of training per se, but also the effects of any other events between 1975 and 1977 whose probabilities are increased conditional on the knowledge that an individual entered training (for example, the probability of unemployment in late 1975). The same is true of θ_0, although the impact of other events for trainees who were

TABLE IX

ESTIMATED TRAINING EFFECTS ON THE PROBABILITY OF EMPLOYMENT:
RANDOM EFFECTS LOGIT MODELS T1 AND T2

	All Trainees	Classroom Trainees	Nonclassroom Trainees
Estimated Increase in Employment Probabilities Due to Training[a]			
1. *Model T1:*			
(a) 1977	.020	.036	.017
(b) 1978	.049	.089	.032
(c) 1979	.071	.130	.056
(d) Average 1977–79	.047	.085	.035
(e) Steady-state[b]	.063	.115	.047
2. *Model T2:*			
(a) 1977	.030	.031	.014
(b) 1978	.063	.084	.034
(c) 1979	.082	.126	.047
(d) Average 1977–79	.058	.080	.032
(e) Steady-state[b]	.063	.114	.051

[a] Difference between fitted probability of employment using estimated training effects, and fitted probability of employment setting training effects to zero.
[b] Difference between steady-state employment probabilities with and without training effects, evaluated at the 1970 year effect.

unemployed in 1975 is less clear. The important point is that the estimates of τ_1 and τ_0, the parameters describing the permanent effects of training, should not be affected. The permanent effects of training should, moreover, dominate in the long run.

As in the previous sections, the results in Table VIII suggest that the training effects are larger for the classroom participants than other CETA participants. The implied increases in the employment probabilities for both groups of trainees are summarized in Table IX. The estimated training effects from both models are quite similar, and are very close to the estimates presented in Table VI from the conditional logit model. Since these effects incorporate both the permanent training effects (τ_0 and τ_1), as well as the once-for-all effects (θ_0 and θ_1), their interpretation is difficult. For comparison rows (1e) and (2e) of Table IX present estimates of the long-run effects of training that depend only on τ_0 and τ_1. These estimates are again larger for the classroom than nonclassroom trainees, although the combined training effect is a relatively substantial 6.3 percent increase in employment probabilities from either model T1 or T2.

The evidence on the goodness-of-fit of models T1 and T2 suggests that the latter model gives a statistically better fit, although the implied training effects are very similar. The estimates of the relative selection probability parameters (ρ_i) and the values and probabilities of the individual effects are presented in Table X. The estimated mass-points (the α_i's) and the estimated fractions of the control and trainee groups of each type are very similar for the two models. The estimated selection ratios for model T1 range from 1.37 (for the classroom

TABLE X

ESTIMATED INDIVIDUAL EFFECTS, TYPE PROBABILITIES, AND
RELATIVE SELECTION RATIOS: RANDOM EFFECTS LOGIT MODELS T1 AND T2

	Description of Estimated Mass Points			
	Type 1	Type 2	Type 3	Type 4
Model T1:				
1. All Trainees				
(a) Log odds of employment (α_i)[a]	−3.70	−.90	.49	1.72
(b) Control weight[b]	.026	.184	.352	.438
(c) Trainee weight[c]	.001	.374	.625	.000
(d) Selection ratio (ρ)[d]	1.74	1.74	1.74	1.74
2. Classroom Trainees				
(a) Log odds of employment (α_i)[a]	−4.50	−1.14	.37	1.86
(b) Control weight[b]	.020	.145	.469	.363
(c) Trainee weight[c]	.000	.395	.605	.000
(d) Selection ratio (ρ)[d]	1.37	1.37	1.37	1.37
3. Nonclassroom Trainees				
(a) Log odds of employment (α_i)[a]	−3.44	−.79	.51	1.43
(b) Control weight[b]	.028	.220	.206	.546
(c) Trainee weight[c]	.003	.394	.603	.000
(d) Selection ratio (ρ)[d]	1.83	1.83	1.83	1.83
Model T1:				
1. All Trainees				
(a) Log odds of employment (α_i)[a]	−3.70	−.86	.85	2.28
(b) Control weight[b]	.026	.202	.571	.201
(c) Trainee weight[c]	.002	.446	.552	.000
(d) Selection ratio (ρ_i)[d]	.63	1.19	4.10	1.17
2. Classroom Trainees				
(a) Log odds of employment (α_i)[a]	−4.70	−1.16	.44	2.04
(b) Control weight[b]	.019	.144	.527	.000
(c) Trainee weight[c]	.000	.396	.604	.000
(d) Selection ratio (ρ_i)[d]	.64	1.07	2.26	1.15
3. Nonclassroom Trainees				
(a) Log odds of employment (α_i)[a]	−3.28	−.69	1.24	1.26
(b) Control weight[b]	.030	.259	.515	.196
(c) Trainee weight[c]	.006	.541	.453	.000
(d) Selection ratio (ρ_i)[d]	.33	1.25	6.72	1.11

[a] Log odds of employment, conditional on unemployment in the previous year and assuming a 1970 year effect.
[b] Estimated fraction of control group of each type.
[c] Estimated fraction of trainee group of each type.
[d] Estimated ratio of the probability of entering training conditional on unemployment in 1975 to the probability of entering training conditional on employment in 1975.

trainees) to 1.74 (for the nonclassroom trainees). The estimated selection ratios for model T2 vary with the value of the individual effects, and suggest that the relative probability of entering training from employment is lower for individuals with higher value of the individual effects. The exception to this pattern is the highest effect type (Type 4 in the table), for whom the estimated selection ratios are close to unity. For both groups of trainees the fraction of the trainee group assigned to the highest type is very small, however,

In summary, our estimates from the nonlinear employment probability models suggest three conclusions regarding the impact of training on the 1976 CETA cohort. First, the estimated effects of CETA participation on subsequent employment probabilities range from 3 to 8 percent, on average, with most of the increase concentrated among classroom trainees. Second, the effects of CETA participation include both transitory effects on 1977 employment status, and permanent effects on post-training transition rates. We have argued that estimates of the former effects are potentially biased by the presence of other unobservable determinants of post-training employment status that are nevertheless correlated with training. Our estimates of the permanent effects of training are on the order of 5 to 10 percentage points, with the larger effects again concentrated among classroom trainees. Third, we find no evidence that the estimated training effects are biased by failure to consider the interaction between individual-specific effects in the probability of employment and individual effects in the probability of entering training.

CONCLUSIONS

The results in this paper suggest that participation in CETA had a small to moderately large positive impact on the post-training employment probabilities of the 1976 cohort of adult male trainees. Our estimates of the effect of training on the average probability of employment during 1977–79 range from 2 to 5 percentage points. The lower range of these estimates is obtained by a comparison of relative pre- and post-training employment probabilities of the CETA trainees and a control group, and also by a comparison between trainees and controls with identical pre-training histories. The upper range of these estimates is obtained by a series of logistic probability models of the employment histories of the trainees and controls.

The methods all point to significantly larger training effects for participants in classroom training programs, as compared to on-the-job programs, although the estimated effects of both types of programs are consistently positive. Many of the on-the-job CETA programs involved little or no formal training, however, and their relatively smaller effect on subsequent employment probabilities is therefore understandable. Since the costs of the classroom training programs were substantially lower than the costs of the nonclassroom programs, our results suggest that the classroom programs were superior in a cost-benefit sense. Assuming that the CETA trainees earned approximately $5800 per year in the post-training period, if employed, and that CETA participation increased the probability of employment in every year after training by 2 to 5 percent, training may have increased participant earnings by $100–$300 per year. This increase compares favorably to the cost of CETA training, which averaged about $1500 per participant in 1976.[36]

[36]Average earnings of the trainees in 1977 and 1978 were $4750 and $5140 (in 1976 dollars) respectively. Assuming an average probability of employment of .85 among the trainees, the average earnings of trainees, conditional on employment, were approximately $5800 in the post-training period. The figure for CETA cost per participant is taken from *Employment and Training Report of the President* (1977, Tables F-2 and F-3, pp. 262–263).

Given that the available data are nonexperimental, there is, of course, ample reason to be cautious in interpreting these results. Nonexperimental methods of program evaluation have recently come under attack for their lack of reliability, and our discussion of the comparisons between observationally identical trainees and controls highlighted many of the difficulties in a nonexperimental evaluation. Nevertheless we have presented several highly overidentified models of employment determination and trainee status that appear to fit the observed data quite well. We have also presented a variety of less heavily parameterized program estimators that give fairly similar estimates of the effectiveness of training. Finally, we have argued that many of the biases that enter an observational study of training effects can be isolated in the once-for-all effects of training on employment status at the end of program participation. Our estimates of the effect of training, abstracting from these one-time effects, give similar but slightly higher estimates of the training effects on employment probabilities.

Department of Economics, Princeton University, Princeton, NJ 08544, U.S.A.
and
Department of Economics, Northwestern University, Evanston, IL 60201, U.S.A.

DATA APPENDIX

In this appendix we describe the sources of the trainee and control group data used in the paper.

1. TRAINEES

The trainee data are taken from the Continuous Longitudinal Manpower Survey (CLMS). The CLMS sample is drawn from participants in programs operated by a stratified random sample of CETA prime sponsors, and contains Social Security Administration earnings records as well as demographic data collected at the date of enrollment. We used CLMS data for the 1976 cohort of trainees, provided to us by SRI International. We included only those members of the sample who were male, 21 years of age or older at enrollment, and who reported enrollment and termination dates between January 1 and December 31, 1976.

2. COMPARISON GROUP

The comparison group data are drawn from the March 1976 Current Population Survey (CPS). Members of the CPS were matched, by their Social Security numbers, to Social Security Administration records of earnings from 1966 to 1979. The comparison group sample was provided to us by SRI International, and includes male CPS members who were 21 years of age or older in March 1976, who reported being in the labor force during the survey week in March, and who reported individual and family earnings during 1975 of less than $20,000 and $30,000, respectively. Further details on the construction of the comparison sample are presented in Dickinson, Johnson, and West (1984, pp. 37–45).

APPENDIX

In this appendix we demonstrate that the minimal sufficient statistic for the fixed effect α_i in the logistic probability model with state dependence is in general the entire vector of observed outcomes for the ith individual. Following equations (6) and (7) of the text, let E_{it} represent an indicator for

whether i is employed in t, and assume that

$$p(E_{it} = 1) = \exp(\alpha_i + \beta_t + \gamma E_{it-1})/(1 + \exp(\alpha_i + \beta_t + \gamma E_{it-1})), \qquad t > 1,$$

$$p(E_{i1} = 1) = \frac{\exp(\alpha_i)(1 + \exp(\alpha_i + \gamma))}{1 + 2\exp(\alpha_i) + \exp(\alpha_i)\exp(\alpha_i + \gamma)}.$$

Dropping i subscripts, the probability of an observed sequence of indicators $E = (E_1, E_2, \ldots E_T)$ is

(A.1) $$p(E, \alpha) = \frac{\exp(\alpha) + \exp(E_1(2\alpha + \gamma))}{1 + 2\exp(\alpha) + \exp(2\alpha + \gamma)} \prod_{t=2}^{T} \frac{\exp(E_t(\alpha + \beta_t + \gamma E_{t-1}))}{1 + \exp(\alpha + \beta_t + \gamma E_{t-1})}.$$

We now construct a minimal sufficient statistic $T(E)$ for α. It can be shown that if $p(E, \alpha_0) > 0$ for all E, then the mapping from E to $p(E, \alpha)/p(E, \alpha_0)$, regarded as a function of α, is a minimal sufficient statistic (for example, Bahadur (1954)). Using (A.1), a minimal sufficient statistic for α is

(A.2) $$T(E)_\alpha = p(E, \alpha)/P(E, 0)$$

$$= \frac{3 + \exp(\gamma)}{1 + 2\exp(\alpha) + \exp(2\alpha + \gamma)} \cdot \frac{\exp(\alpha) + \exp(E_1(2\alpha + \gamma))}{1 + \exp(E_1\gamma)}$$

$$\cdot \frac{\exp\left(\alpha \sum_{t=2}^{T} E_t\right)}{\displaystyle\prod_{t=2}^{T} \frac{1 + \exp(\alpha + \beta_t + \gamma E_{t-1})}{1 + \exp(\beta_t + \gamma E_{t-1})}}.$$

In case $\gamma = 0$, (A.2) can be written as

$$T(E)_\alpha = \frac{\exp\left(\alpha \sum_{t=1}^{T} E_t\right)}{\displaystyle\prod_{t=1}^{T} \frac{1 + \exp(\alpha + \beta_t)}{1 + \exp(\beta_t)}},$$

(defining $\beta_1 = 0$), which is a function only of the total number of years of employment $\sum_{t=1}^{T} E_t$.

In the general case, if there is a minimally sufficient statistic of dimension smaller than the data vector $(E_1, \ldots E_T)$, then there exist $(E_1, \ldots E_T)$ and $(E_1', \ldots E_T')$ such that $T(E)_\alpha = T(E')_\alpha$ for all real α and $E \neq E'$. Let $S = \sum_{t=1}^{T} E_t$ and $S' = \sum_{t=1}^{T} E_t'$. Making use of the fact that

$$\frac{\exp(\alpha) + \exp(E_1(2\alpha + \gamma))}{1 + \exp(E_1\gamma)} = \frac{\exp(\alpha E_1)(1 + \exp(\alpha + \gamma E_1))}{1 + \exp(\gamma E_1)},$$

equality of $T(E)_\alpha$ and $T(E')_\alpha$ implies

(A.3) $$e^{\alpha S}(1 + e^\alpha \exp(\gamma E_1))(1 + \exp(\gamma E_1')) \prod_{t=2}^{T} (1 + \exp(\beta_t + \gamma E_{t-1}))$$

$$\times (1 + e^\alpha \exp(\beta_t + \gamma E_{t-1}'))$$

$$= e^{\alpha S'}(1 + e^\alpha \exp(\gamma E_1'))(1 + \exp(\gamma E_1))$$

$$\times \prod_{t=2}^{T} (1 + \exp(\beta_t + \gamma E_{t-1}'))(1 + e^\alpha \exp(\beta_t + \gamma E_{t-1})).$$

This is a polynomial expression in e^α. If the right-hand and left-hand sides of (A.3) are equal for all

real α, then the polynomials must have the same degree, implying $S = S'$. Similarly, the polynomials must have the same constant terms, implying

$$
\left(1 + \exp\left(\gamma E_1'\right)\right) \prod_{t=2}^{T} \left(1 + \exp\left(\beta_t + \gamma E_{t-1}\right)\right)
$$

$$
= \left(1 + \exp\left(\gamma E_1\right)\right) \prod_{t=2}^{T} \left(1 + \exp\left(\beta_t + \gamma E_{t-1}'\right)\right).
$$

Simplifying (A.3), we have

$$
\left(1 + e^{\alpha} \exp\left(\gamma E_1\right)\right) \prod_{t=2}^{T} \left(1 + e^{\alpha} \exp\left(\beta_t + \gamma E_{t-1}'\right)\right)
$$

$$
= \left(1 + e^{\alpha} \exp\left(\gamma E_1'\right)\right) \prod_{t=2}^{T} \left(1 + e^{\alpha} \exp\left(\beta_t + \gamma E_{t-1}\right)\right)
$$

for all real α. Thus we must have equality of the sets $\{\gamma E_1, \beta_2 + \gamma E_1', \beta_3 + \gamma E_2', \ldots, \beta_T + \gamma E_{T-1}'\}$ and $\{\gamma E_1', \beta_2 + \gamma E_1, \beta_3 + \gamma E_2, \ldots, \beta_T + \gamma E_{T-1}\}$. Provided that $\gamma \neq 0$ and that the year effects are not all zero, this requires $E_t = E_t'$ $(t = 1, \ldots, T-1)$ for all but exceptional values of the structural parameters. Thus (A.3) implies that $E = E'$, so there can be no nontrivial sufficient statistic.

REFERENCES

ANDERSON, E.B. (1973): *Conditional Inference and Models for Measuring*. Copenhagen: Mentalhygejnisk Forlag.
ASHENFELTER, O. (1975): "The Effect of Manpower Training on Earnings: Preliminary Results," *Proceedings of the Twenty-Seventh Annual Winter Meetings of the Industrial Relations Research Association*, ed. by J. Stern and B. Dennis. Madison: Industrial Relations Research Association.
―――― (1978): "Estimating the Effect of Training Programs on Earnings," *Review of Economics and Statistics*, 60, 47–57.
ASHENFELTER, O. AND D. CARD (1985): "Using the Longitudinal Structure of Earnings to Estimate the Effect of Training Programs," *Review of Economics and Statistics*, 67, 648–660.
BAHADUR, R.R. (1954): "Sufficiency and Statistical Decision Functions," *Annals of Mathematical Statistics*, 25, 423–462.
BASSI, L. (1983): "The Effect of CETA on the Post-Program Earnings of Participants," *The Journal of Human Resources*, 18, 539–556.
CHAMBERLAIN, G. (1980): "Analysis of Covariance with Qualitative Data," *Review of Economic Studies*, 47, 225–238.
DICKINSON, K., T. JOHNSON, AND R. WEST (1984): "An Analysis of the Impact of CETA Programs on Participants' Earnings," Final Report prepared for the U.S. Department of Labor Employment and Training Administration. Menlo Park, CA: SRI International.
Employment and Training Report of the President (1976 and 1977). Washington, D.C.: Government Printing Office.
GOLDBERGER, A. (1972): "Selection Bias in Evaluating Treatment Effects," Discussion Paper No. 123-172, Institute for Research on Poverty, University of Wisconsin.
HECKMAN, J. (1979): "Sample Selection Bias as a Specification Error," *Econometrica*, 47, 153–161.
HECKMAN, J., AND B. SINGER (1984): "A Method for Minimizing the Impact of Distributional Assumptions in Econometric Models for Duration Data," *Econometrica*, 52, 271–320.
HECKMAN, J., AND R. ROBB (1986): "Alternative Methods for Solving the Problem of Selection Bias in Evaluating the Impact of Treatments on Outcomes," in *Drawing Inferences from Self Selected Samples*, ed. by H. Wainer. New York: Springer Verlag.
KAITZ, H. (1979): "Potential Use of Markov Process Models to Determine Program Impact," in *Research in Labor Economics*, Supplement 1, ed. by Farrell Bloch. Greenwich, CT: JAI Press.
KENDALL, M., AND A. STUART (1973): *The Advanced Theory of Statistics*. New York: Hafner Publishing Co.
KIEFER, J., AND J. WOLFOWITZ (1956): "Consistency of the Maximum Likelihood Estimator in the Presence of Infinitely Many Incidental Parameters," *Annals of Mathematical Statistics*, 27, 887–906.

KIEFER, N. (1979): *The Economic Benefits of Four Employment and Training Programs.* New York: Garland Publishing, Inc.

LINDSAY, B. (1983a): "The Geometry of Mixture Likelihoods, Part 1," *Annals of Statistics*, 11, 86–94.

——— (1983b): "The Geometry of Mixture Likelihoods, Part II," *Annals of Statistics*, 11, 783–792.

LAIRD, N. (1978): "Nonparametric Maximum Likelihood Estimation of a Mixing Distribution," *Journal of the American Statistical Association*, 73, 805–811.

LALONDE, R. (1986): "Evaluating the Econometric Evaluations of Training Programs with Experimental Data," *American Economic Review*, 76, 604–620.

LALONDE, R., AND R. MAYNARD (1987): "How Precise are Evaluations of Employment and Training Programs: Evidence from a Field Experiment," *Evaluation Studies*, 11, 428–451.

RAASCH, G. (1960): *Probabilistic Models for Some Intelligence and Attainment Tests.* Copenhagen: Denmarks Paedogogiske Institute.

ROSENBAUM, P., AND D. RUBIN (1985a): "Reducing Bias in Observational Studies Using Subclassification on the Propensity Score," *Journal of the American Statistical Association*, 79, 516–524.

——— (1985b): "Constructing a Control Group Using Multivariate Matched Sampling Methods that Incorporate the Propensity Score," *The American Statistician*, 39, 33–38.

Social Security Bulletin Annual Statistical Supplement (1983). Washington, D.C.: Government Printing Office.

LINKED CITATIONS
- Page 1 of 4 -

[Footnotes]

[3] **Evaluating the Econometric Evaluations of Training Programs with Experimental Data**
Robert J. LaLonde
The American Economic Review, Vol. 76, No. 4. (Sep., 1986), pp. 604-620.
Stable URL:
http://links.jstor.org/sici?sici=0002-8282%28198609%2976%3A4%3C604%3AETEEOT%3E2.0.CO%3B2-P

[12] **Using the Longitudinal Structure of Earnings to Estimate the Effect of Training Programs**
Orley Ashenfelter; David Card
The Review of Economics and Statistics, Vol. 67, No. 4. (Nov., 1985), pp. 648-660.
Stable URL:
http://links.jstor.org/sici?sici=0034-6535%28198511%2967%3A4%3C648%3AUTLSOE%3E2.0.CO%3B2-N

[16] **Evaluating the Econometric Evaluations of Training Programs with Experimental Data**
Robert J. LaLonde
The American Economic Review, Vol. 76, No. 4. (Sep., 1986), pp. 604-620.
Stable URL:
http://links.jstor.org/sici?sici=0002-8282%28198609%2976%3A4%3C604%3AETEEOT%3E2.0.CO%3B2-P

[17] **Estimating the Effect of Training Programs on Earnings**
Orley Ashenfelter
The Review of Economics and Statistics, Vol. 60, No. 1. (Feb., 1978), pp. 47-57.
Stable URL:
http://links.jstor.org/sici?sici=0034-6535%28197802%2960%3A1%3C47%3AETEOTP%3E2.0.CO%3B2-K

NOTE: *The reference numbering from the original has been maintained in this citation list.*

LINKED CITATIONS

- Page 2 of 4 -

[24] **Evaluating the Econometric Evaluations of Training Programs with Experimental Data**
Robert J. LaLonde
The American Economic Review, Vol. 76, No. 4. (Sep., 1986), pp. 604-620.
Stable URL:
http://links.jstor.org/sici?sici=0002-8282%28198609%2976%3A4%3C604%3AETEEOT%3E2.0.CO%3B2-P

[25] **Reducing Bias in Observational Studies Using Subclassification on the Propensity Score**
Paul R. Rosenbaum; Donald B. Rubin
Journal of the American Statistical Association, Vol. 79, No. 387. (Sep., 1984), pp. 516-524.
Stable URL:
http://links.jstor.org/sici?sici=0162-1459%28198409%2979%3A387%3C516%3ARBIOSU%3E2.0.CO%3B2-6

[26] **Analysis of Covariance with Qualitative Data**
Gary Chamberlain
The Review of Economic Studies, Vol. 47, No. 1, Econometrics Issue. (Jan., 1980), pp. 225-238.
Stable URL:
http://links.jstor.org/sici?sici=0034-6527%28198001%2947%3A1%3C225%3AAOCWQD%3E2.0.CO%3B2-L

[32] **Analysis of Covariance with Qualitative Data**
Gary Chamberlain
The Review of Economic Studies, Vol. 47, No. 1, Econometrics Issue. (Jan., 1980), pp. 225-238.
Stable URL:
http://links.jstor.org/sici?sici=0034-6527%28198001%2947%3A1%3C225%3AAOCWQD%3E2.0.CO%3B2-L

References

Estimating the Effect of Training Programs on Earnings
Orley Ashenfelter
The Review of Economics and Statistics, Vol. 60, No. 1. (Feb., 1978), pp. 47-57.
Stable URL:
http://links.jstor.org/sici?sici=0034-6535%28197802%2960%3A1%3C47%3AETEOTP%3E2.0.CO%3B2-K

NOTE: *The reference numbering from the original has been maintained in this citation list.*

LINKED CITATIONS
- Page 3 of 4 -

Using the Longitudinal Structure of Earnings to Estimate the Effect of Training Programs
Orley Ashenfelter; David Card
The Review of Economics and Statistics, Vol. 67, No. 4. (Nov., 1985), pp. 648-660.
Stable URL:
http://links.jstor.org/sici?sici=0034-6535%28198511%2967%3A4%3C648%3AUTLSOE%3E2.0.CO%3B2-N

Sufficiency and Statistical Decision Functions
R. R. Bahadur
The Annals of Mathematical Statistics, Vol. 25, No. 3. (Sep., 1954), pp. 423-462.
Stable URL:
http://links.jstor.org/sici?sici=0003-4851%28195409%2925%3A3%3C423%3ASASDF%3E2.0.CO%3B2-V

Analysis of Covariance with Qualitative Data
Gary Chamberlain
The Review of Economic Studies, Vol. 47, No. 1, Econometrics Issue. (Jan., 1980), pp. 225-238.
Stable URL:
http://links.jstor.org/sici?sici=0034-6527%28198001%2947%3A1%3C225%3AAOCWQD%3E2.0.CO%3B2-L

Sample Selection Bias as a Specification Error
James J. Heckman
Econometrica, Vol. 47, No. 1. (Jan., 1979), pp. 153-161.
Stable URL:
http://links.jstor.org/sici?sici=0012-9682%28197901%2947%3A1%3C153%3ASSBAAS%3E2.0.CO%3B2-J

A Method for Minimizing the Impact of Distributional Assumptions in Econometric Models for Duration Data
J. Heckman; B. Singer
Econometrica, Vol. 52, No. 2. (Mar., 1984), pp. 271-320.
Stable URL:
http://links.jstor.org/sici?sici=0012-9682%28198403%2952%3A2%3C271%3AAMFMTI%3E2.0.CO%3B2-U

Consistency of the Maximum Likelihood Estimator in the Presence of Infinitely Many Incidental Parameters
J. Kiefer; J. Wolfowitz
The Annals of Mathematical Statistics, Vol. 27, No. 4. (Dec., 1956), pp. 887-906.
Stable URL:
http://links.jstor.org/sici?sici=0003-4851%28195612%2927%3A4%3C887%3ACOTMLE%3E2.0.CO%3B2-Z

NOTE: *The reference numbering from the original has been maintained in this citation list.*

LINKED CITATIONS

- Page 4 of 4 -

The Geometry of Mixture Likelihoods: A General Theory

Bruce G. Lindsay
The Annals of Statistics, Vol. 11, No. 1. (Mar., 1983), pp. 86-94.
Stable URL:
http://links.jstor.org/sici?sici=0090-5364%28198303%2911%3A1%3C86%3ATGOMLA%3E2.0.CO%3B2-4

The Geometry of Mixture Likelihoods, Part II: The Exponential Family

Bruce G. Lindsay
The Annals of Statistics, Vol. 11, No. 3. (Sep., 1983), pp. 783-792.
Stable URL:
http://links.jstor.org/sici?sici=0090-5364%28198309%2911%3A3%3C783%3ATGOMLP%3E2.0.CO%3B2-B

Nonparametric Maximum Likelihood Estimation of a Mixing Distribution

Nan Laird
Journal of the American Statistical Association, Vol. 73, No. 364. (Dec., 1978), pp. 805-811.
Stable URL:
http://links.jstor.org/sici?sici=0162-1459%28197812%2973%3A364%3C805%3ANMLEOA%3E2.0.CO%3B2-J

Evaluating the Econometric Evaluations of Training Programs with Experimental Data

Robert J. LaLonde
The American Economic Review, Vol. 76, No. 4. (Sep., 1986), pp. 604-620.
Stable URL:
http://links.jstor.org/sici?sici=0002-8282%28198609%2976%3A4%3C604%3AETEEOT%3E2.0.CO%3B2-P

Reducing Bias in Observational Studies Using Subclassification on the Propensity Score

Paul R. Rosenbaum; Donald B. Rubin
Journal of the American Statistical Association, Vol. 79, No. 387. (Sep., 1984), pp. 516-524.
Stable URL:
http://links.jstor.org/sici?sici=0162-1459%28198409%2979%3A387%3C516%3ARBIOSU%3E2.0.CO%3B2-6

NOTE: *The reference numbering from the original has been maintained in this citation list.*