DOCUMENT RESUME

ED 137 330 TM 006 130

AUTHOR	Jacobson, Louis S.
TITLE	The Use of Longitudinal Data to Assess the Impact of Manpower Training on Earnings.
INSTITUTION	Public Research Inst., Arlington, Va.
SPONS AGENCY	Department of Labor, Washington, D.C.
REPORT NO	PRI-73-2
PUB DATE	20 Jul 73
CONTRACT	DOL-L-72-86
NOTE	49p.
EDRS PRICE	MF-$0.83 HC-$2.06 Plus Postage.
DESCRIPTORS	Comparative Analysis; Data Analysis; Employment Patterns; *Federal Programs; *Income; *Longitudinal Studies; *Manpower Development; Manpower Needs; Manpower Utilization; Mathematical Models; Multiple Regression Analysis; Statistical Analysis; *Trainees; *Training
IDENTIFIERS	Manpower Development and Training Act; Social Security Data

ABSTRACT

Research using Social Security data to measure the effects of government manpower training programs on the earnings of participants is discussed. Previous studies employed a longitudinal set of Social Security summary earnings records of over 50,000 individuals who participated in Manpower Development and Training Act (MDTA) institutional training in 1964 and 70,000 individuals randomly selected from the same file. This study used Social Security's Continuous Work History Sample (CWHS) to demonstrate that these prior studies have substantially underestimated the impact of training on earnings. The studies being criticized employed a model of income determination that did not take into account the fact that trainees were induced to join the program because they were having difficulty finding or holding adequate jobs. The basic technique used in this paper was to compare actual post-training earnings of trainees to the earnings of a control group whose earnings potential at the time of initiation of training was identical to that of the participants. In order to compare the two groups, a model of income determination was specified and estimated using regression analysis. Since it was not possible to identify trainees in the CWHS or to obtain a reliable measure of unemployment and loss of job tenure in the summary earnings records, the estimation procedure was based on knowledge of the simple correlations among the variables. (Author/MV)

PRI 73-2

The Use of Longitudinal Data to Assess the Impact of Manpower Training on Earnings

Louis S. Jacobson

PUBLIC RESEARCH INSTITUTE
a division of the Center for Naval Analyses

PRI STUDIES

Brown, George, and Silverman, Lester. The Retail Price of Heroin: Estimation and Applications. Public Research Institute report PRI 73-1. May 1973.

Jacobson, Louis. The Use of Longitudinal Data To Assess the Input of Manpower Training on Earnings. Public Research Institute report PRI 73-2. July 1973.

Silverman, Lester. The Determinants of Daily Emergency Admission to Hospitals. Public Research Institute report PRI 73-3. July 1973.

Holen, Arlene, and Horowitz, Stanley. The Effect of Unemployment Insurance and Eligibility Enforcement on Unemployment. Public Research Institute report PRI 74-1. April 1974.

Silverman, Lester, et al. Undiagnosed Bacterial Meningitis in Vermont Children. Public Research Institute report PRI 74-2. October 1974.

Jondrow, James; Devine, Eugene; Jacobson, Louis; Katz, Arnold; O'Neill, David. Removing Restrictions on Imports of Steel. Public Research Institute report PRI 75-2. May 1975.

PRI 73-2 60 730200.00

The Use of Longitudinal Data to Assess the Impact of Manpower Training on Earnings

Louis S. Jacobson

20 July 1973

Work was supported by the Labor Department on contract L–72–85

PUBLIC RESEARCH INSTITUTE
1401 Wilson Boulevard
Arlington, Virginia 22209

ACKNOWLEDGEMENT

I would like to acknowledge the assistance of George Delehanty of Northwestern University for making the data available, Jim Tuccy of Vogelback Computer Center, Northwestern University for programming assistance, Orley Ashenfelter of the Department of Labor for sharing the results of his work and providing constructive criticism of this report, and Paul Feldman, Director of the Public Research Institute, for helping to edit this report. Any errors are solely the responsibility of the author.

ABSTRACT

This paper discusses research using Social Security data to measure the effect of government manpower training programs on the earnings of participants. Previous studies employed a longitudinal set of Social Security summary earnings records of over 50,000 individuals who participated in MDTA institutional training in 1964 and 70,000 individuals randomly selected from the same file. This study uses Social Security's Continuous Work History Sample (CWHS) to demonstrate that these prior studies have substantially underestimated the impact of training on earnings.

The studies being criticized employed a model of income determination that did not take into account the fact that trainees were induced to join the program because they were having difficulty finding or holding adequate jobs.

The basic technique used in this paper is to compare actual post-training earnings of trainees to the earnings of a control group whose earnings potential at the time of initiation of training was identical to that of the participants. In order to compare the two groups, a model of income determination is specified and estimated using regression analysis.

Because it was not possible to identify trainees in the CWHS or to obtain a reliable measure of unemployment and loss of job tenure in the summary earnings records, the estimation procedure was based on knowledge of the simple correlations among the variables.

TABLE OF CONTENTS

LIST OF TABLES

8

LIST OF FIGURES

Introduction

This report discusses the use of Social Security data to measure the effect of Manpower Development and Training Act (MDTA) programs on the earnings of participants. Two studies conducted at the U.S. Department of Labor (DOL), have produced the counter-intuitive finding that MDTA training has a substantial negative effect on subsequent earnings. This paper demonstrates that these findings do not constitute a valid assessment of the effectiveness of MDTA training. It shows that training significantly increases the earnings of participants above what they could expect without training. Finally it recommends procedures that can be used by the DOL to produce a more accurate assessment of the returns to training.

The first part of this paper examines the basic problems that must be solved to analyze correctly the impact of MDTA training on earnings. It shows that the models of income determination used in evaluating MDTA training are crucial to the proper solution of these problems; an appropriate model must contain a variable to measure the influence of factors which induced the trainees to join the MDTA program rather than to remain in the private labor market. The findings of the earlier studies are shown to be in error because they used an inappropriate model to measure the relative earnings of MDTA trainees. The theoretical differences between the original model and several re-specified models are examined.

In the second part of this paper, the impact of MDTA training on earnings is calculated using re-specified models. Initially, the calculation is made using data that differ slightly from those in the DOL sample. An appropriate model is then developed that uses the DOL data. The findings based on these models indicate that the MDTA program has been successful in raising the relative earnings of participants.

Finally, several additional modifications of the basic model are made in order to produce even more accurate assessments of the effectiveness of MDTA training. The overall conclusions of this study are, first, that the MDTA program has enabled participants to raise their earnings above what they would have obtained without training and, second, that although the studies cited reached invalid conclusions, Social Security data can be used efficiently to obtain an accurate assessment of the impact of manpower programs on future earnings.

The Problem: Selection of a Control Group

The basic problem in determining the impact of MDTA training on future earnings is to measure the difference between the actual post-training earnings of the trainees and the amount these individuals would have earned had they not participated in the program. To do this requires comparing the trainees' earnings to the earnings of different individuals whose earnings potential is identical to that of the trainees at the time of initiation of training.

The ideal control group would be developed by randomly selecting the trainees and the controls from a pool of qualified applicants. Unfortunately, this technique has proven extremely difficult to implement.

The alternative used in most studies is to select a control group randomly from the labor market. A statistical matching procedure is then developed to create artificially a measure of earnings for a group of individuals who did not participate in the program but whose earning potential was identical to that of the trainees. This procedure must be able to compensate for variations between the trainees and controls in the distribution of specific characteristics that affect earnings. There are several ways to do this, of which regression analysis is probably the most efficient. Use of any analytic method, however, requires the specification of a model of income determination that includes all the independent variables influencing earnings which may be distributed differently between the two groups. By far the greatest source of difficulty in developing a useable model is the limited availability of data that include the relevant characteristics.* Because the data constraint is crucial in the creation of a useable model, the nature of the data used in the MDTA studies is described below.

The Social Security Administration provided information to the Department of Labor about the age, sex, and race of over 50,000 individuals who participated in MDTA training in 1964, the amount of earnings each year on which Social Security tax was collected,**

*This problem is avoided by using the "ideal" control group, since presumably, any characteristic that affects income is distributed identically in the control and trainee groups.

**Obviously, no earnings are reported for employment not covered by Social Security nor are earnings reported over the taxable limit ($4800 from 1958-1965). However, total yearly earnings above the maximum are extrapolated.

and the number of quarters each year in which earnings exceeded $50 for five pre-training years, 1958-1962, and five post-training years, 1965-1969. Data for the year 1964 were omitted in order to exclude the training period itself. Data for 1963 were also omitted, since some participants in 1964 were also in the program during 1963.

Similar data were provided for individuals who participated in MDTA programs in 1968, but only one year of post-training data is currently available for this group. This analysis focuses primarily on those who terminated enrollment in the MDTA Institutional program in 1964. To provide a comparison group for trainees in both 1964 and 1968, the records of more than 70,000 individuals who did not receive MDTA training were randomly selected from Social Security files, and the same variables were reported for this group.

All models underlying research conducted with these data specify that, for members of each of four race-sex cohorts, post-training earnings were directly related to the individual's age, pre-training earning characteristics, and participation in MDTA training. The general model used initially by DOL researchers is presented in equation (1).

$$Y_t = f(A, Y_{1958}, \ldots, Y_{1962}, Q_{1958}, \ldots, Q_{1962}, T_{1964}) \qquad (1)$$
$$(t = 1965, \ldots, 1969)$$

where A = age

Y_i = earnings in year i

Q_i = number of quarters earnings exceed $50, year i

T = training dummy, 1 = trainees

Two specific formulations derived from this general model are described below.

David Farber examined a model presented in equation (2).*

$$Y_t = g(A, \sum_{j=1958}^{1962} E_j, T_{1964})^{**} \qquad (2)$$

*Farber, David, An Analysis of Change in Earnings of Participants in Manpower Training Programs, DOL Internal Report, 1972.

**The comparison procedure also utilized an earnings pattern measure but it has been omitted from the model since it did not significantly affect the results.

where

$$E_j = \frac{Y_j}{Q_j}$$

The comparison measure of post-training earnings was established by the following procedure: Within each race-sex cohort the proportion of trainees falling into each of fifty unique cells based on ten age categories and five average pre-training earnings categories was calculated. The average post-training earnings of the members of the comparison group falling into each of the same fifty cells was ascertained. Finally, the proportion of trainees in each cell was multiplied by the average post-training earnings of the controls in the same cell and the products were summed for all fifty cells. This sum was the average post-training earnings of a group who did not participate in MDTA training but whose specific age and average pre-training earnings were identical to those of the trainees.

Orley Ashenfelter examined the model specified in equation (3)* using the same data.

$$Y_t = a_0 + a_1 A + \sum_{i=1958}^{1962} b_{E_i} E_i + cT_{1964} \qquad (3)$$

The comparison of post-training earnings between trainees and controls entailed a straightforward use of regression analysis.

These models, relying solely on age and pre-training earnings to predict the future earnings of individuals in a given race-sex cohort, might appear to include far too few variables to take into account the variation in the distribution of all possible variables that influence earnings. Many factors such as education, IQ, and work experience have been used successfully to account for the variation in individual earnings. The use of these factors is based on human capital theory in which an individual's earnings are assumed to be a complex function of ability and training. However, human capital theory also implies that an individual's earnings are an excellent measure of the sum total of all these diverse factors. Thus, one might expect that two groups matched with respect

*Ashenfelter, Orley, Analysis of Social Security Data to Detect Possible Biases in Trainee-Control Comparisons, DOL Memo, December 1972.

to race and sex, with identical distributions of age and earnings, have almost identical distributions of specific characteristics which influence earnings. Therefore, explicitly including these human capital characteristics in the above model may not change the estimate of the effect of training on earnings (coefficient c).*

Regardless of whether earnings or other measures are used to estimate an individual's stock of human capital at a specific point in time, his subsequent actions can change his stock. In many cases these changes will be randomly distributed between two "matched" populations. However, trainees and controls examined by DOL researchers experienced systematically different changes in their level of human capital. As a result omission of data in the year immediately before entering the program (1963) led to a biased estimate of the level of human capital of the trainees and to an erroneous conclusion concerning the effectiveness of MDTA training. This report explains the nature of the bias and estimates its magnitude.

Empirical studies of income determination that have used the human capital approach have relied on cross-sectional analyses rather than following individual earning profiles over time. Thus, there is little direct evidence in the literature for assessing the validity of any of the models presented. The results of Farber's study, presented in table I, are based on the general model in equation (1). They indicate that the MDTA program was, in most cases, not merely ineffectual but deleterious to the relative earnings of participants. When identical data were used with Ashenfelter's model, the results were similar.

The Effect of Unemployment on Future Earnings

A possible interpretation of these findings is not that training destroys a worker's productivity, but that the human capital that would be lost by leaving the labor force to undertake MDTA training is greater than increments that would be gained through the training. However, the MDTA program was not designed to raise earnings by supplementing the employment experience of individuals in the labor market who already hold jobs. Rather, it was intended to provide compensatory training for specific individuals who were having trouble finding adequate employment in the private labor market and

*Recent work by Orley Ashenfelter has confirmed this point with regard to introducing trainees' education into the model.

TABLE I

AVERAGE YEARLY POST-TRAINING EARNINGS OF MDTA INSTITUTIONAL COMPLETERS AND THE CONTROL GROUPS SELECTED BY FARBER FROM SOCIAL SECURITY DATA: 1964, 1968

	White male	White female	Black male	Black female
1964*				
Average Post-Training Earnings of:				
Trainees	$3932	$2112	$3130	$1976
Controls	4132	2044	3130	1788
Change in Earnings Due to Training:				
Amount	-200	68	0	188
Percent	-5.0	3.3	0	10.5
1968**				
Average Post-Training Earnings of:				
Trainees	$3308	$2068	$2440	$1804
Controls	4036	2320	3148	2092
Change in Earnings Due to Training:				
Amount	-728	-252	-708	-288
Percent	-18.0	-10.9	-22.5	-13.8

*Post-training earnings in 1964 is the average of 1965-1969 earnings.
**Post-training earnings in 1968 reflects earnings in 1969 only.

otherwise would likely have been involuntarily unemployed. This might be due either to permanent displacement from former occupations or to an inability to hold a job because of insufficient training. That the institution chosen to approve candidates for MDTA training is the Employment Service and that the individuals volunteering are almost all unemployed* reinforce this view.

The appropriate measure of the impact of the program on the trainees' future earnings is the present discounted value of the differences in trainees' earnings and the earnings of others equally likely to have volunteered for training.**

The models examined by Farber and Ashenfelter have failed to include explicitly an indicator of the forces leading to entrance into the program such as a variable measuring involuntary unemployment directly before entering the program. The omission of earnings for 1963 from their model precluded the specification of an adequate proxy. As a consequence, these models may have mis-specified the true model.

The central question about the models discussed is: Has the estimated effect of MDTA training been significantly biased by omitting a measure of unemployment? As previously indicated, any variable influencing earnings can legitimately be omitted from the first three models only when the variable is equally distributed within cohorts of the comparison and trainee groups after these cohorts are matched on the basis of the variables included in the relevant model.

With respect to the distribution of unemployment, it has already been pointed out that MDTA records show almost 100% of the trainees are unemployed when applying for entrance to the program. Statistics derived from Social Security data,*** however, indicate that fewer than 50% of the individuals in the comparison group experienced any unemployment during the entire year preceding entrance. Because of the size of this disparity, it is very unlikely that any cohort in the comparison group will have as high an

*The unemployment status of MDTA enrollees is reported in the Manpower Report of the President, March 1973, p. 231.

**The studies discussed in this report neither examined the amount of earnings foregone during training nor calculated the value of the program to society. This report is limited to determining the accuracy with which the studies measured the post-training earnings of similar individuals with and without MDTA training.

***This calculation was made from a set of Social Security data available to the author. For further discussion of the calculation, see appendix D.

incidence of unemployment as the corresponding trainee cohort. Since unemployment is demonstrated in this paper to reduce subsequent earnings severely, it is evident that in the absence of training the trainee group would be expected to have significantly lower earnings than the comparison group. The extent to which the estimated earnings of trainees is biased downward is determined by two factors: the average difference in earnings between those who were unemployed and those who were not, and the average difference in the incidence of unemployment between controls and trainees.

Equation (4) presents a re-specified model that includes a measure of unemployment in 1963 -- (U_{1963}).

$$Y_t = \tilde{a}_0 + \tilde{a}_1 A + \sum_{i=1958}^{1962} \tilde{b}_{Y_i} Y_i + \tilde{c} T_{1964} + \tilde{d} U_{1963} \tag{4}$$

The difference between the original model represented by equation (3) and the re-specified model represented by equation (4) can be depicted graphically.

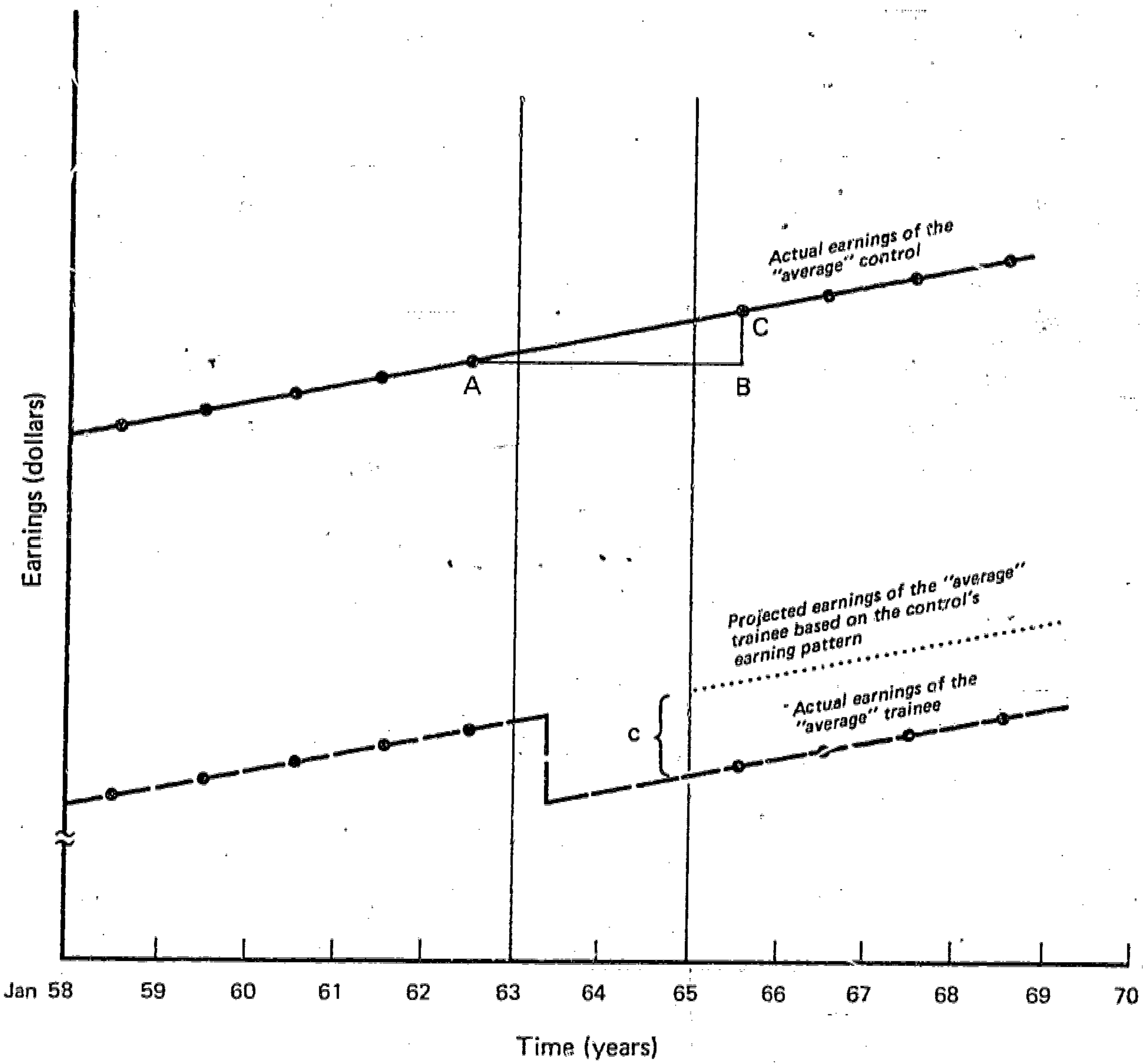
A regression based on equation (3) predicts an individual's post-training earnings, given his age and pre-training earnings, so that over all individuals in the sample, the sum of squared differences between actual earnings and predicted earnings is minimized. The regression "line," therefore, can be said to give an unbiased estimate of the actual earnings of the "average" individual in the sample.

The solid line in figure 1 depicts the time path of earnings for the "average" individual in the DOL sample. Since average earnings increase more or less uniformly, * earnings in the post-training period are projected to increase proportionately at the rate indicated by the ratio of line segment BC to AB. This value is related to the value of the coefficient b_{E62} in equation (3). **

The effect of MDTA training on post-training earnings is normally estimated by examining the difference between actual earnings and predicted earnings of trainees based on the above projection. The dashed line in figure 1 depicts the actual time path of earnings of the "average" trainee. This line falls below the solid line because the

*The discussion utilizes this and other assumptions in order to simplify the analysis and aid comprehension. The assumptions are reasonably close to reality.

**Equation (4) was estimated both as shown and with $E_i = \frac{i}{2}$ substituted for Y_i. The two formulations yielded very similar results.



Each heavy dot represents a single data point. Note: No data is included between January 1963 and December 1964.

FIG. 1: ILLUSTRATION OF THE CALCULATION OF THE IMPACT OF MDTA TRAINING ON EARNINGS BASED ON REGRESSION EQUATION 3

average earnings of trainees are considerably lower than the average earnings of all individuals in the DOL sample. The discontinuous drop in earnings during 1963 is due to the unemployment of the average trainee. The dotted line represents the projected earnings in the post-training period based on the above regression procedure. The vertical difference (c) between the two lines equals the negative impact of training on earnings estimated by the DOL studies and is equivalent to coefficient c in equation (3).

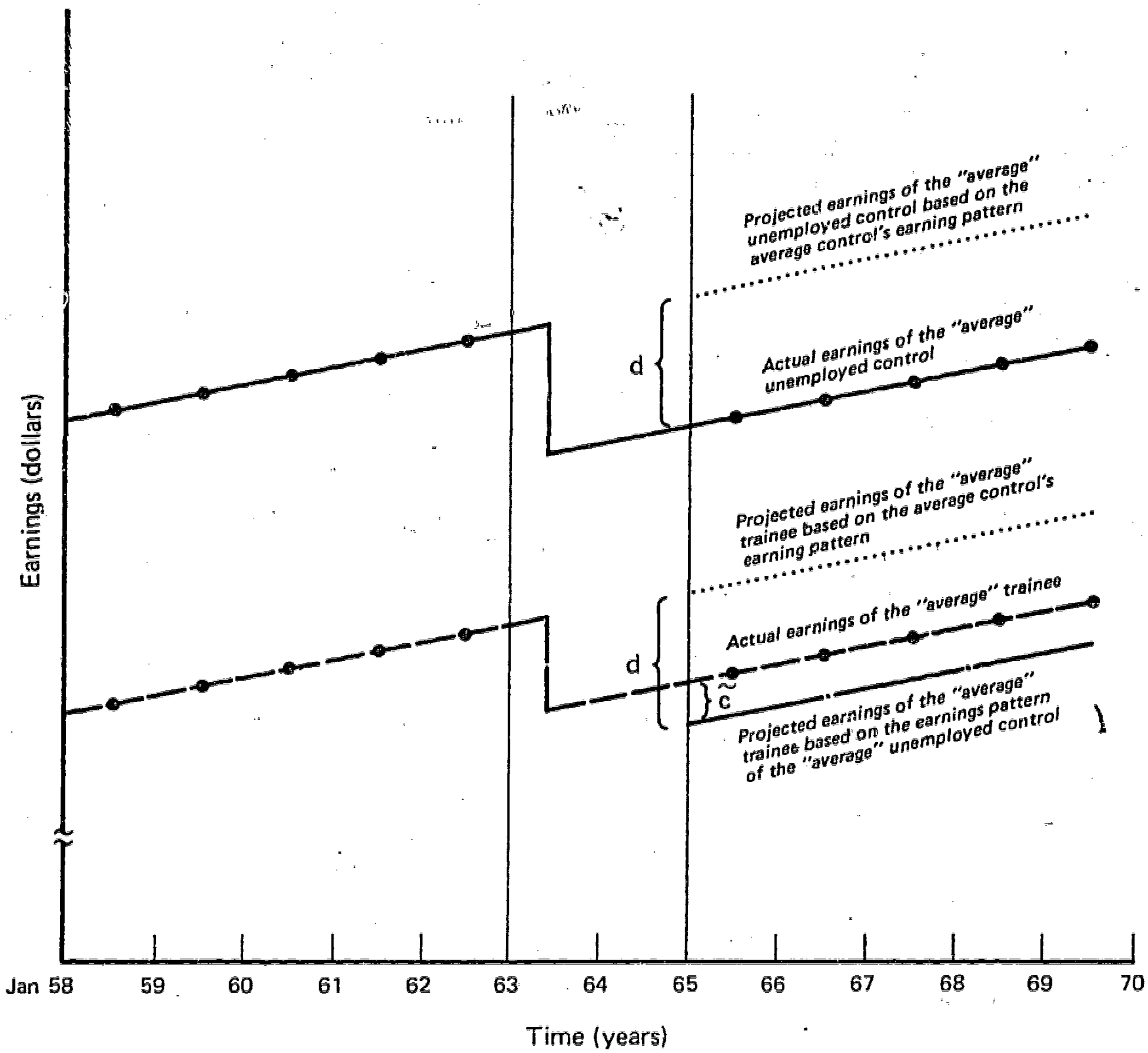
This formulation is a valid comparison between the trainees and the average individual in the Social Security sample. The criticism resolved by the re-specified model is that the comparison group is inappropriate because trainees have all experienced unemployment. Since the trainees are atypical, they should be compared to similar individuals, not to "average" individuals. Figure 2 shows how such a comparison can be made.

The solid line in figure 2 represents the time path of earnings for the average individual unemployed in 1963. The dotted line indicates the projected earnings of such an individual based on equation (3). The difference (d) between the actual earnings and the projected earnings represent the influence of unemployment on earnings. This value is related to the coefficient d in equation (4). The dashed line, which is identical to the dashed line in figure 1, indicates the actual earnings of trainees. The dash-dot line indicates the projected value of trainees' earnings based on the assumption that all trainees are unemployed before entering the program. The difference ($\tilde{c}$) between the two lines is the effect of MDTA training on earnings. The value is equal to coefficient $\tilde{c}$ in equation (4).

Estimate of the Bias Due to Mis-specification

These "common sense" concepts are given precise mathematical formulation below. The bias (B) in the estimate of the amount MDTA training influences earnings is equal to the difference between the estimated coefficient (c in equation (3)) and the true coefficient reflecting the effect of MDTA training on future earnings ($\tilde{c}$ in equation (4)). The relationship between these coefficients is expressed in equation (5). The derivation is presented in appendix A.

$$B = c - \tilde{c} = u_T \tilde{d} \qquad (5)$$



Each heavy dot represents a single data point. Note: No data is included between January 1963 and December 1964.

FIG. 2: ILLUSTRATION OF THE DIFFERENCE BETWEEN THE CALCULATION OF THE IMPACT OF MDTA TRAINING BASED ON REGRESSION EQUATION 3 AND REGRESSION EQUATION 4 WHICH INCLUDES PRE-TRAINING UNEMPLOYMENT

where:

($\tilde{d}$) is the regression coefficient appearing in equation (4) and is a measure of the impact of unemployment on future earnings, and

(u_T) is the regression coefficient shown in equation (6) and is an estimate of the difference in the probability of being unemployed in 1963, between trainees and controls.

$$U_{1963} = u_o + u_A A + \sum_{i=1957}^{1962} u_{Y_i} Y_i + u_T T_{1964} \tag{6}$$

As discussed above, it is very likely that u_T is positive since most trainees are unemployed, and $\tilde{d}$ is negative because unemployment lowers future earnings. Thus, the product of these coefficients is negative. As expected, this formulation indicates that the influence of training on future earnings has been biased downward.

The precise magnitude of the bias could be revealed by estimating the coefficients $\tilde{d}$ and u_T in equations (4) and (6). However, Social Security data do not include a specific measure of unemployment. A proxy measure of unemployment has been developed by using a set of Social Security data, available to the author, which supplements the variables in the data with additional information. The unemployment measure is based on detection of the dip in quarterly earnings that accompanies unemployment.*

Unfortunately, the new data set does not identify MDTA trainees so it is not possible to estimate directly either $\tilde{d}$ or u_T. However, it is possible to estimate $\tilde{d}$ indirectly using this new data set.

The details of the estimation procedure appear in appendix B. The estimate of $\tilde{d}$, reported in table II, line 4, shows that the impact of unemployment on earnings is relatively large. The average earnings of unemployed individuals were more than 20% lower than the earnings of individuals unaffected by unemployment.

*For a full discussion of the data used and the variable created see appendix D.

TABLE II

ESTIMATES* OF THE BIAS IN CONTROL GROUP'S EARNINGS DUE TO OMISSION OF UNEMPLOYMENT FROM THE BASIC MODEL

		White male	White female	Black male	Black female
1.	Estimated Earnings of the Control Group	$2820	$2446	$2404	$1708
2.	Estimated Percent Change in Earnings Due to MDTA Training (from table I)	-5	3.3	0	10.5
3.	Estimated Dollar Change in Earnings Due to MDTA Training (Coefficient c)(Line 2 x Line 1)	-$141	79	0	$180
4.	Estimated Impact of Unemployment in 1963 on Earnings (Coefficient $\tilde{d}$)	-547	-446	-305	-272
5.	Estimated Difference in the Probability of Being Unemployed Between Trainees and Controls (Coefficient u_T)	.44	.46	.39	.24
6.	Estimated Bias Due to Omitting Unemployment ($u_T \cdot d$)	-$240	-$205	-$119	-$ 65
7.	Corrected Dollar Estimate of Change in Earnings Due to MDTA Training (Line 3 - Line 6)	$ 99	$284	$119	$245
8.	Corrected Percent Change in Earnings Due to MDTA Training [Line 7 ÷ (Line 1 + Line 3)]	3.7	11.2	5.0	13.0

*These estimates are based on the methodology described in appendices B and C. All dollar figures are based on the earnings of the controls in 1963.

It is also possible to estimate u_T . The estimation depends upon a very complex procedure that utilizes the new data set and specific knowledge about the DOL sample. The details of this procedure appear in appendix C. The results are reported in table II, line 5.

The estimated bias, $u_T\tilde{d}$, is presented in table II, line 6. The new, more appropriate estimate of the impact of MDTA training on earnings is revealed by subtracting the estimate of the bias due to omitting unemployment from the original estimate of the impact of MDTA training. This result is reported in dollar terms in table II, line 7, and in percentage terms on line 8. It is clear the MDTA program is considerably more effective than was indicated in previous work, since the effects are all positive, ranging from $100 to $245.

An Alternative Method of Estimating the Effect of Training

Because of the numerous assumptions necessary to estimate the bias, a more accurate estimate of MDTA effectiveness could be achieved by introducing an unemployment variable directly into the DOL data. As previously indicated, this can only be done by using a proxy measure of unemployment. The proxy discussed above is based on quarterly earnings. It cannot be used with the Social Security data available to the Department of Labor because only annual earnings are included in the DOL data. Rather than constructing a dummy variable based on observing reductions in quarterly earnings, a process which provides a measure of the incidence of unemployment, it is possible to use earnings in 1963 as a substitute.* Since it is likely that most of the variation in earnings is due to fluctuations in duration of employment, it is even possible that the earnings variable more closely measures the effect of actual unemployment than does the initial unemployment proxy.**

*To be sure that earnings in 1963 is not biased, it is necessary either to delete from the sample those 1964 completers who entered the program in 1963 or to divide the earnings of these individuals by the proportion of the year they were not in the MDTA program. These steps are needed since, for these individuals, earnings in 1963 will be inordinately low. Failure to exclude these trainees would lead to an under-estimate of their subsequent earnings, biasing upward the estimates of the effect of training on earnings.

**However, the unemployment variable can detect individuals whose yearly earnings are nearly constant but who are unemployed for part of every year. This pattern may be typical of a large number of low income individuals in the sample.

Equation (7) is a modification of the re-specified model shown in equation (4). In this formulation, earnings in 1963 is used as a proxy for unemployment.

$$Y_{1965} = \alpha_0 + \alpha_1 A \sum_{i=1958}^{1962} \beta_{Y_i} Y_i + \gamma T + \delta Y_{1963} \tag{7}$$

The difference between the estimated impact of MDTA training on future earnings in the original model and in the above model is described in equation (8),

$$B_1 = c - \gamma \tag{8}$$

It is a simple matter to determine this difference since both c and γ can be estimated directly using the DOL sample.

These estimates are shown in table III, lines 2 and 3.

The foregoing discussion implies that if Y_{1963} is an adequate proxy for U_{1963}, the bias measured by using the unemployment proxy based on quarterly earnings (B) and the bias measured by using earnings in 1963 (B_1) should be roughly equal. Table III, lines 3 and 4, presents this comparison.

The considerably higher estimate of the bias using the unemployment variable (U_{1963}) may indicate that Y_{1963} is not as sensitive a measure of unemployment as U_{1963} and the model based on equation (7) probably underestimates the effectiveness of MDTA training. *

This possibility is supported by the fact that when U_{1963} is introduced in equation (7) (with T deleted) the estimated coefficient δ remains practically unchanged while the coefficient associated with U_{1963} is highly significant and equal to -$125.00. This indicates that the effectiveness of training is underestimated by approximately $55.

The Effect of Loss of Job Tenure on Future Earnings

The use of the unemployment proxies discussed above may not be sufficient to remove all the bias due to omitting from the original models the factors associated with entering

*The higher estimate of B relative to B_1 may be due, in part, to the use of a different subsample in the estimation procedure. However, it is likely the use of identical samples would only accentuate this difference. See appendix B for a further discussion of this point.

TABLE III

COMPARISON OF THE BIAS IN THE ESTIMATES OF THE IMPACT OF MDTA TRAINING ON EARNINGS CALCULATED USING ALTERNATIVE PROXIES FOR UNEMPLOYMENT IN 1963

		White male
1.	Estimate of Bias Using Quarterly Earnings Dip Proxy (Table II, Line 6)	-$240
2.	Original Estimate of Impact of MDTA Training on Earnings* (c of Equation (3))	247
3.	New Estimate of the Impact of MDTA Training on Earnings (γ of Equation (7))	409
4.	Estimate of Bias Using Earnings in 1963 as Unemployment Proxy ($c - \gamma$)	- 162

*This figure differs from Farber's calculation because individuals with zero earnings in 1964 were excluded from Farber's control group. This exclusion biased downward the estimate of the effectiveness of MDTA training. (Estimates of c and γ are currently available for white males only.)

training. Both of the variables used detect the loss of earnings associated with involuntary unemployment which in turn closely relates to trouble finding or holding an adequate job. However, individuals who suffer an equal loss of earnings will not all suffer an equal loss in their stock of human capital. A loss of earnings will also occur in cases of voluntary withdrawal from the labor force, temporary layoffs, or in cases in which heavy overtime is followed by a return to a normal work week. These actions are usually not indicative of employment trouble.

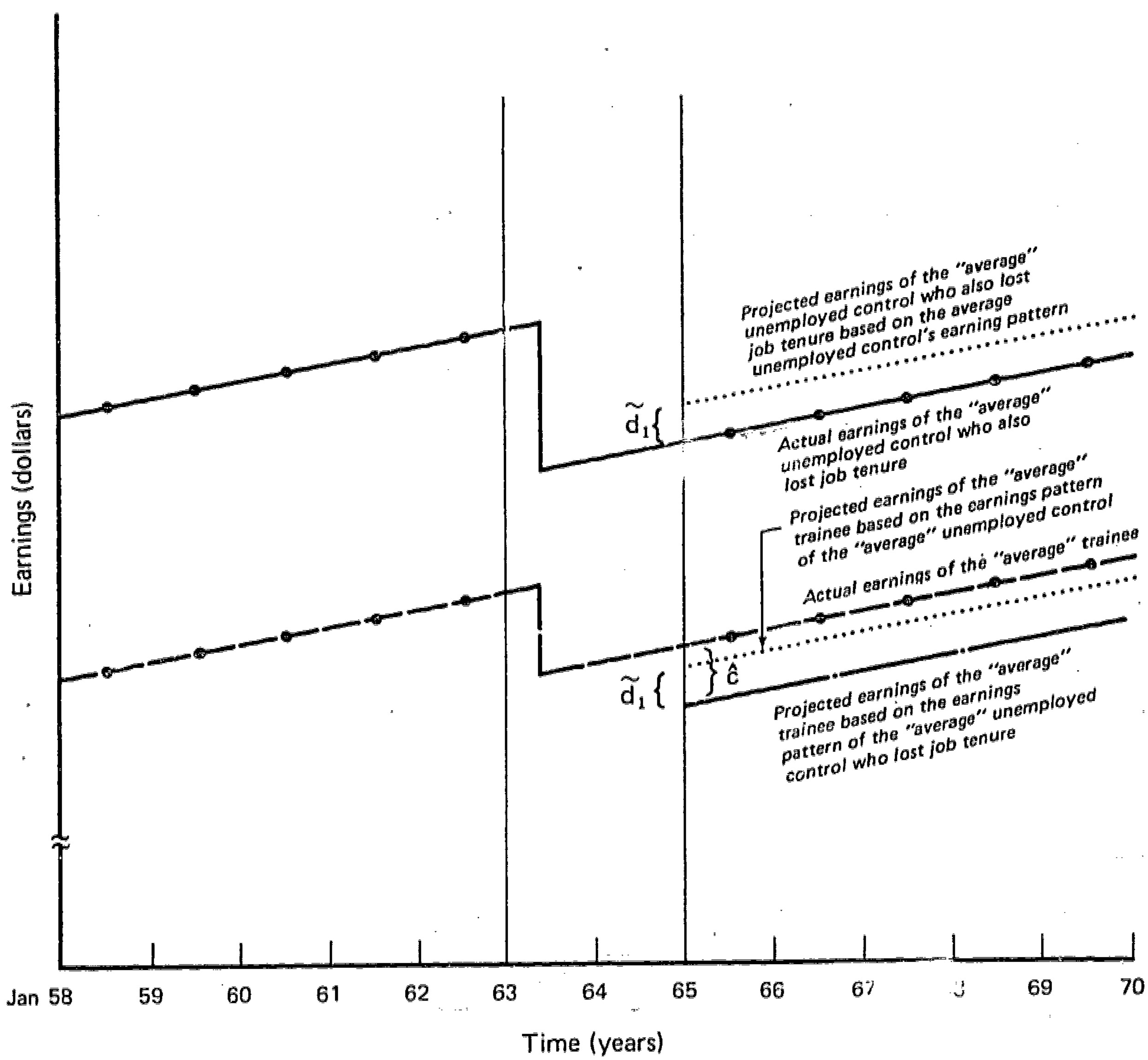
Importantly, voluntary withdrawal and temporary layoffs, from which an individual returns to the same firm or at least the same occupation, may have a considerably less depressing effect on future earnings than would involuntary unemployment after which an individual must change firms and often change occupations.

In practical terms, an individual who changes jobs must, at the very least, learn the routine of his new firm and frequently learn to perform new tasks. He will not have the benefits of seniority and will have to prove himself to his employer in order to secure wage increases or promotions. This loss of "firm-specific" training and experience is likely to be proportionally greater for low income individuals with few transferable skills.

The unemployment measure used in equation (4) does not distinguish between individuals who changed jobs and thereby lost their specific training, and those who went back to the same job. Because practically all MDTA trainees are involuntarily unemployed and must change firms, while the control group very likely includes a far lower proportion of similar individuals, the omission of a loss of job tenure variable probably biases downward the estimate of the impact of MDTA training on future earnings, even after unemployment is taken into account.

The foregoing discussion can be illustrated graphically. The solid line in figure 3 represents the real time-path of earnings for the "average" individual who was unemployed in 1963 and suffered a loss of "specific training" human capital. The dotted line represents the projected time path of earnings for this individual based on equation (4).* The vertical difference ($\tilde{d}_1$) between actual earnings and projected earnings represents the bias due to omitting the loss of specific training from equation (4). The dashed line

*The dotted line is equivalent to the solid line shown in figure 2.



Each heavy dot represents a single data point. Note: No data is included between January 1963 and December 1964.

FIG. 3: ILLUSTRATION OF THE DIFFERENCE BETWEEN THE CALCULATION OF THE IMPACT OF MDTA TRAINING BASED ON REGRESSION EQUATION 4 AND REGRESSION EQUATION 9 WHICH INCLUDES LOSS OF JOB TENURE

is the earnings of the average trainee.* Assuming all trainees lose job tenure, the new earnings projection is represented as the dash-dot line. The impact of MDTA training on future earnings is $\hat{c}$. As shown in the figure, $\hat{c}$ is greater than $\tilde{d}_1$.

Thus, a measure of the loss of job tenure should be introduced into equation (4) in addition to a measure of unemployment. The new formulation is presented in equation (9).

$$Y_t = \hat{a}_0 + \hat{a}_1 A + \sum_{i=1957}^{1962} \hat{b}_{Y_i} Y_i + \hat{c}\, T_{1964} + \hat{d}_0 U_{1963} + \hat{d}_1 LN_{1963} \tag{9}$$

where LN = a measure of job tenure (longevity).** The bias due to omitting both unemployment and job tenure is shown in equation (10).

$$B_2 = c - \hat{c} = u_T \hat{d}_0 + \ell n_T \hat{d}_1 \quad . \tag{10}$$

u_T is precisely the regression coefficient indicated in equation (6) and ℓn_T is the coefficient obtained by re-estimating equation (6) after replacing U by LN as the dependent variable. The size of bias B_2 can be calculated using the same methods used to determine bias B. The results of these calculations are shown on table IV. The interested reader is again referred to appendices B and C for a discussion of the methodology.

Th size of the additional bias eliminated by including a proxy for the loss of job tenure into the model already containing a proxy for unemployment is presented in table IV, line 9. The added influence of the loss of job tenure is important for all groups but comparatively smaller for whites than blacks.***

This may be due in part to the fact, noted earlier, that not all job changes lead to a loss of human capital. Many individuals voluntarily change jobs to apply their skills for a new employer at an increased salary. The present formulation does not efficiently

*This dashed line is equivalent to the dashed line in figures 1 and 2.

**The loss of job tenure variable is based on detecting a change of employers between two years. For further details see appendix D.

***The positive bias for black females and other anomolous results for this group are probably a result of a different set of factors than those that affect other groups. Possibly the key factor is the disproportionate occurrence of multiple job holders in the black-female cohort.

TABLE IV

ESTIMATES OF THE BIAS IN CONTROL GROUP EARNINGS DUE TO OMITTING UNEMPLOYMENT AND LOSS OF JOB TENURE FROM THE BASIC MODEL

	White male	White female	Black male	Black female
1. Estimated change in earnings due to MDTA training (based on Farber's calculation)				
a. $(From table II, line 3)	-$141	$79	$0	$180
b. %(From table II, line 2)	-5%	3.3%	0%	10.5%
2. Estimated impact of unemployment in 1963 on earnings (coefficient $\hat{d}_0$)	-510	-424	-249	-290
3. Estimated differential probability of being unemployed between controls and trainees (Coefficient u_T)	0.44	0.46	0.39	0.24
4. Estimated impact of loss of job tenure in 1963 on earnings (Coefficient $\hat{d}_1$)	-99	-109	-169	110
5. Estimated differential probability of losing job tenure between controls and trainees (Coefficient ℓn_T)	0.66	0.77	0.64	0.75
6. Estimated Bias due to omitting unemployment and loss of job tenure				
a. $u_T \cdot \hat{d}_0$	-224	-195	-97	-70
b. $\ell n_T \cdot \hat{d}_1$	-65	-84	-108	82
c. $u_T \hat{d}_0 + \ell n_T \cdot \hat{d}_1$	-289	-279	-205	12
7. Corrected Estimate of Change in Earnings Due to MDTA Training				
a. $ (Line 1a - Line 6c)	148	358	205	168
b. %	5.2%	14.6%	8.5%	9.8%
8. Estimated Bias due to omitting unemployment only (table II, line 6)	-240	-205	-119	-65
9. Change in bias due to including loss or job tenure in addition to unemployment				
a. $ (Line 6c - Line 8)	49	74	86	-77
b. % (Line 9a ÷ Line 8)	-20.4%	-36%	-72.3%	118%

distingush between this type of voluntary job change and an involuntary job change which is typically associated with decreased earnings. Thus, the higher the proportion of voluntary job changers in the sample, and the greater the difference in earning between voluntary and involuntary job changers, the more likely it is that the job tenure coefficient will be positive. The impact of the loss of job tenure may be considerably underestimated because the distinction between voluntary and involuntary job change is omitted.

The Effect of Introducing a Four-Way Classification of Unemployment and Loss of Job Tenure

This possible bias could be greatly reduced, if not eliminated, by creating dummy variables that indicate simultaneously whether or not an individual is unemployed and whether or not he loses job tenure. These variables would allow distinctions to be made between those individuals who suffer involuntary unemployment and change jobs with a subsequent loss of human capital and individuals who voluntarily change jobs. Although both groups would show a "loss" of job tenure, presumably the latter group will suffer little unemployment. The variables would also distinguish these individuals from those who suffer temporary unemployment and return to their former jobs and those who neither change jobs nor suffer unemployment. This four-way classification is described in table V. A re-specified model including these dummy variables is shown in equation (11):

$$Y_t = a_0 + a_1 A + \sum_{i=1958}^{1962} b_{y_i} Y_i + cT_{1964} + d_0 TU + d_1 JA + d_2 IU \tag{11}$$

where

TU = temporary unemployment, $U = 1$, $LN = 0$

JA = job advancement $U=0$, $LN = 1$

IU = involuntary unemployment and loss of job tenure $U = 1$, $LN = 1$

Examination of table VI indicates that at least 25% of all job changers suffer no unemployment. Contrary to expectations, white females are slightly more likely than males to change jobs without suffering unemployment. Table VII shows that the relative difference in earnings between job changers who also suffer unemployment and those

TABLE V

THE RELATION OF UNEMPLOYMENT AND LOSS OF JOB TENURE TO LABOR MARKET ACTIVITY

	Stable earnings (No unemployment) U = 0	Dip in earnings (Unemployment) U = 1
No employer change LN = 0	I Steady employment with a single firm (0)	II Withdrawal from the labor force or temporary layoff (-)
Employer change LN = 1	III Job advancement (+)	IV Involuntary unemployment and loss of job tenure (-)*

The signs in parentheses indicate the expected influence on future earnings relative to steady employment.

*Both withdrawal and involuntary unemployment lead to an earnings decline relative to steady employment. The decline associated with involuntary unemployment is more severe a priori.

TABLE VI

THE PROPORTION OF EACH COHORT IN EACH
OF THE FOUR UNEMPLOYMENT - LOSS OF JOB TENURE CATEGORIES

	White male	White female	Black male	Black female
1. Percent steadily employed	45.4	50.7	42.0	39.7
2. Percent unemployed without loss of job tenure	21.5	26.6	22.7	36.8
3. Percent changed jobs without unemployment	8.3	8.0	9.1	5.8
4. Percent changed jobs and unemployed	24.8	14.7	26.2	17.7

TABLE VII

THE AVERAGE IMPACT ON FUTURE EARNINGS OF UNEMPLOYMENT
OR LOSS OF JOB TENURE RELATIVE TO STEADY EMPLOYMENT

	White male	White female	Black male	Black female
5. Unemployment without loss of job tenure	-$478	-$436	-$235	-$336
6. Job change without unemployment	-$ 32*	-$138	-$142	-$ 51*
7. Job change accompanied by unemployment	-$617	-$524	-$420	-$159

*Not significantly different from zero at the 10% level.

who do not is very substantial. This difference is particularly dramatic for white males. These findings tend to confirm the hypothesis that the impact of the loss of job tenure on the earning of trainees is substantially underestimated by the model specified in equation (9). It strongly suggests that the most accurate appraisal of the MDTA program would be obtained by using as controls only individuals known to have been unemployed and to have changed jobs.

Summary and Conclusions

This study has clearly demonstrated that mis-specification of the basic model of income determination has led to an incorrect estimate of the impact of MDTA training on the subsequent annual earnings of all groups of participants except black females. Rather than decreasing earnings by 5%, training increases earnings by 5%.

After David Farber's calculations were corrected for the estimated bias, estimates of subsequent earnings of participants in the 1964 MDTA Institutional program were substantially higher than what they could have expected without training. If, in addition to the correction for bias, individuals with zero earnings in 1964 are included in the control group, the impact of MDTA training proves even more substantial.

Most importantly, this report has demonstrated that Social Security data can be used effectively to evaluate manpower training programs. An accurately specified model of income determination can be estimated with these data because the earnings of an individual in a given race-sex-age cohort are an adequate measure of current human capital and the proxy variables allowing a four-way classification of unemployment and job tenure provide adequate measures of subsequent changes in human capital.

Since unemployment and job tenure measures cannot be constructed from the data sample currently available at the DOL, earnings in the year immediately before training must be substituted. Although this modification of the initial model leads to a substantial reduction in the bias, the impact of MDTA training may still be underestimated by as much as 50% for white males. (Precise estimates of the remaining bias are not available for other groups but it is likely that the differential is considerably smaller.)

It would be highly desirable to base future evaluation of any manpower training program on the 1% CWHS. However, the sample size is too small to produce the records of a sufficient number of trainees and the cost of collecting additional data would probably be prohibitive. An alternative that is comparatively easy to implement would be to preselect the control group from the 1% CWHS so that the proportion of individuals who fall into each of the four unemployment-job tenure categories immediately before the initiation of training is identical for controls and trainees.*

Use of this procedure requires knowledge of the proportion of trainees in each unemployment-job tenure category. Although this proportion is reasonably clear for the MDTA institutional program, it would be advantageous, for future work, to collect this information for all programs as part of the Management Information System. Alternatively, it would be possible to calculate it directly from a small sample of trainee records from the CWHS. This procedure even if used only once would provide an excellent means for checking the accuracy of the job tenure and unemployment variables used in this report against questionnaire data. It would also provide an opportunity for more accurately examining the differences among alternative models.

Although substantial progress has been made in developing the methodology for using Social Security data to analyze manpower programs, additional work may still prove useful. A study is being planned to determine the impact of local labor market conditions on income determination, the effect of utilizing different sub-samples on the estimated coefficients in the various models of income determination, and the value of improving the accuracy of the variables utilized in the models.

*This procedure would require deletion from the sample of all individuals with earnings above the taxable limit. Very few trainees, if any, would be included given the current limit. Therefore, this inclusion will only tend to make the controls more similar to trainees.

APPENDIX A - MIS-SPECIFICATION ANALYSIS*

Let equation (1) represent a "true" standard regression model in matrix notation.

$$Y = X\tilde{\beta} + u \tag{1}$$

Y = (Nx1) column vector of the dependent variable

X = (NxT) data matrix of the independent variable

$\tilde{\beta}$ = (Tx1) column vector of regression coefficients

u = (Tx1) column vector of residuals

N = number of observations in the data

T = number of independent variables in the model

Assume equation (1) can be partitioned such that

$$X = (X_1, X_2), \quad X_1 :(NxT_1), \quad X_2 :(NxT_2) \tag{2a}$$

$$\tilde{\beta} = \begin{pmatrix} \tilde{\beta}_1 \\ \tilde{\beta}_2 \end{pmatrix} \qquad \tilde{\beta}_1 :(T_1x1), \ \tilde{\beta}_2 :(T_2x1) \tag{2b}$$

where X_1 is the data matrix of variables used in equation (3) below
X_2 is the data matrix of omitted variables

then let equation (3) represent the estimated regression model

$$Y = X_1\hat{\beta}_1 + e \tag{3}$$

$$\hat{\beta}_1 = (X_1'X_1)^{-1}X_1'Y \text{ is the OLS estimator for equation (3)} \tag{4}$$

The bias in the vector of estimated coefficients relative to the "true" estimate is estimated by the following manipulation:

Substituting equation (1) into equation (4):

$$\hat{\beta}_1 = (X_1'X_1)^{-1}X_1'(X\tilde{\beta} + u) \tag{5}$$

*For further discussion see Theil, H., "Specification Errors and the Estimation of Economic Relationship," Review of the International Statistical Institute, Vol. 25, pp. 41-51, 1957.

Since the expected value of β is unaffected by the vector of residual $(E(u) = 0$ by definition).

$$E(\hat{\beta}_1) = (X_1'X_1)^{-1}X_1'X\tilde{\beta} \tag{6}$$

Substituting equations (2a) and (2b) into equation (5):

$$E(\hat{\beta}_1) = (X_1'X_1)^{-1}X_1'(X_1, X_2)\begin{pmatrix}\tilde{\beta}_1 \\ \tilde{\beta}_2\end{pmatrix} \tag{7}$$

Carrying out the matrix multiplication:

$$E(\hat{\beta}_1) = (X_1'X_1)^{-1}\, X_1'X_1\tilde{\beta}_1 + (X_1'X_1)^{-1}X_1X_2\tilde{\beta}_2 \tag{8}$$

and simplifying produces the final result

$$E(\hat{\beta}_1) - \tilde{\beta}_1 = Z\tilde{\beta}_2 \quad \text{where } Z = (X_1'X_1)^{-1}X_1'X_2 \tag{9}$$

or in terms of individual coefficients

$$E(\hat{b}_i) - \tilde{b}_i = \sum_{j=1}^{T-T_1} z_{ij}\tilde{b}_{T_1+j} \tag{10}$$

Thus, the bias in the ith estimated coefficient is the sum of the ijth element of the Z matrix (which is the OLS regression coefficient of i on j) times the T+ith coefficient of the true model.

If there is only a single omitted variable as assumed in the main paper $(T-T_1=1)$

$$E(b_i) - \tilde{b}_i = z_T\tilde{d}_T$$

APPENDIX B

CALCULATION OF $\tilde{d}$ - THE IMPACT OF UNEMPLOYMENT ON EARNINGS

This appendix examines the methodology used to estimate coefficient $\tilde{d}$ of equation (4) (page 8) in the main body of the paper. Equation (4) is reproduced below and relabeled equation (1B).

$$Y_t = \tilde{a}_0 + \tilde{a}_1 A - \sum_{i=1958}^{1962} \tilde{b}_{Y_i} Y_i + \tilde{c}T + \tilde{d}U_{1963} \qquad (1B)$$

As discussed in the text, the data set which contains the most suitable unemployment proxy does not identify MDTA trainees. Consequently it is not possible to estimate $\tilde{d}$ directly. However, it is possible to estimate $\tilde{d}$ indirectly using this new data set. The omission of the variable (T) from equation (1B) will lead to a conservative estimate of the coefficient of $\tilde{d}$. Because the new data do not encompass the precise set of years covered by the DOL data, equation 1B must also be modified by estimating earnings in 1963 instead of 1964 and by using only two years of "pre-training" earnings in the model instead of five. The effect of these changes on the estimate of the influence of unemployment on earnings was examined with supplemental regressions and found to be unimportant. The modified model is presented in equation (2B).

$$Y_{1963} = \tilde{\tilde{a}}_0 + \tilde{\tilde{a}}_1 A + \sum_{1959}^{1960} \tilde{\tilde{b}}_{Y_i} Y_i + \tilde{\tilde{d}}U_{1961} \qquad (2B)$$

Coefficient $\tilde{\tilde{d}}$ above is equal to $\tilde{d}$ of equation (1B). *

These coefficients (and all subsequent coefficients) were estimated using a subset of the sample. The subset included individuals who in 1959 were between the ages of 23 and 53, had earnings less than $4,800 in each of the pre-training years and had earnings greater than $50 in at least one quarter each year. The use of this subset greatly facilitated the preparation and interpretation of measures of difficulty in finding or holding jobs and eliminated some possible sources of bias in these new variables.

*This formulation was also used to calculate $\hat{d}_0$ and $\hat{d}_1$ in equation (10) by introducing LN_{1961} as well as U_{1961} into equation (2B).

Orley Ashenfelter's work indicated that the estimated impact of MDTA training on white males is \$150 higher if the subset described above is used in conjunction with the model specified in equation (7) instead of using the entire sample and the identical model. Most of this difference is probably attributable to the fact that a disproportionate number of white males showing no job change during the five-year period were excluded from the sample because their earnings were greater than the taxable limit. The differential between the sub-sample and full data set is probably smaller for other groups since fewer individuals are excluded for that reason; still, with each cohort, individuals in the sub-sample more closely resemble trainees than do individuals in the full sample. Thus, it would be reasonable to use the sub-sample as the comparison group and it is almost certainly true that the estimate of $\approx{d}$ using the subset produces a conservative estimate of $\tilde{d}$ based on the entire sample. It is the estimate of coefficient $\approx{d}$ that is reported in table II, line 4, page 13.

APPENDIX C

CALCULATION OF u_T - THE DIFFERENTIAL IN THE PROBABILITY OF BEING UNEMPLOYED BETWEEN TRAINEES AND CONTROLS

This appendix examines the methodology used to estimate coefficient u_T of equation (6) on page 12 and in ℓn_T* of equation (9), page 19 in the main body of the paper. The former equation is reproduced below and relabeled equation 1C.

$$U_{1963} = u_0 + u_A A + \sum_{i=1957}^{1962} u_{Y_i} Y_i + u_T T_{1964} \quad (1C)$$

As with the estimation of $\widetilde{d}$, the determination of u_T is made indirectly using the subsample of the data available to the author and described in appendix B.

The estimation of coefficient u_T is based on expressing u_T in terms of simple correlation coefficients and standard deviations. The relationship is presented in equation (2C).

$$u_T = \frac{r_{UT} - r_{UZ} \cdot r_{ZT}}{1 - r_{ZT}^2} \cdot \frac{S_U}{S_T} \quad (2C)$$

where:

- r = The simple correlation coefficient between the subscripted variables.
- S = The standard deviation of the subscripted variable.
- Z = The linear combination of a constant term, and the variables A , Y_{59} , Y_{60} in equation (1B) which maximize r_{UZ}

Fortunately, sufficient information is available to estimate upper and lower bounds for the magnitude of u_T with some confidence.

The linear combination which maximizes the correlation between U and Z is determined by regressing U against the variables included in Z . Assuming that the relationship is roughly equal for controls and trainees, the correlation coefficient (r_{UZ})

*As discussed in the text, the methodology for calculating ℓn_T is identical to the procedure outlined for u_T in this appendix.

specified in equation (2C) is equal to the square root of the coefficient of multiple correlation associated with the regression presented in equation (1B).

Because both the unemployment and training variables can be treated as random variables with the probability p that the variable equals one (and, therefore, indicates either unemployment or trainee status) and the probability $1-p$ that the variable equals zero, the simple correlation between the two variables and their standard deviations can be expressed in terms of these probabilities. Equation (3C) presents this relationship.

$$r_{UT} = \frac{p_{U/T} \cdot p_T - p_U \cdot p_T}{S_U \cdot S_T} \tag{3C}$$

where:

$$p_U = p_{U/T=1} \cdot p_T + p_{U/T=0} \cdot 1-p_T$$

$$S_i = \sqrt{p_i \cdot (1-p_i)} \quad \text{where } i = T, U$$

$p_{U/T}$ = the conditional probability that $U = 1$ given $T = 1$.

Thus, the simple correlation between U and T is a function of only three parameters - p_T, $p_{U/T=1}$, and $p_{U/T=0}$.

The probability that an individual selected at random from the DOL sample is a trainee is equal to the proportion of trainees in the sample. The proportion varies slightly for different race-sex groups but, overall, it is approximately equal to 5/12 or .42 since there are 50,000 trainees out of a total of 120,000 individuals in the sample. Thus, ($p_T = 0.42$). The probability that a trainee is unemployed is known from MDTA statistics and very close to one. Thus, $p_{U/T=1} = 1$. The probability of a control being unemployed is calculated from the social security data available to the author.

The correlation coefficient between Z and T cannot be calculated precisely with the information available. Therefore, an upper and lower bound for r_{ZT} was used in equation (1C).

The lower bound of u_T can be determined by calculating the value of $r_{Z/T}$ that minimizes equation (1C) given the value of the other variables. The calculation of the

upper bound for u_T was based on the assumption $r_{ZT} = 0$* . Estimates of the parameters necessary to calculate u_T are presented in table C-I along with values of u_T derived from those parameters. Similar information concerning the calculation of ℓn_T is provided in table C-II.

*There is no finite upper bound to the value of u_T since as $r_{ZT} \rightarrow 1$, $u_T \rightarrow \infty$. However, it is very unlikely that the value of r_{ZT} is high enough so that this value of u_T exceeds the value of u_T when r_{ZT} equals zero.

TABLE C-I

ESTIMATES OF THE PARAMETERS USED TO CALCULATE u_T - THE DIFFERENTIAL PROBABILITY OF BEING UNEMPLOYED BETWEEN TRAINEES AND CONTROLS

	White male	White female	Black male	Black female
1. Maximum correlation between U and Z - (r_{UZ})	.44	.50	.48	.56
2. Probability that an individual* is a trainee - (p_T)	.42	.42	.42	.42
3. Probability that a trainee* is unemployed in 1963 - ($p_{U/T=I}$)	1.0	1.0	1.0	1.0
4. Probability that a control* is unemployed in 1963 - ($p_{U/T=0}$)	.27	.24	.28	.32
5. Probability that an individual is unemployed - (p_U)	.69	.66	.70	.74
6. Standard deviation of the trainee probability - (S_T) (variable 2)	.49	.49	.49	.49
7. Standard deviation of the unemployment probability (variable 5) - (S_U)	.46	.47	.46	.44
8. Correlation between U and T - (r_{UT})	.57	.61	.56	.50
9. Correlation between Z and T that minimizes u_T - (r_{ZT})	.46	.52	.63	**
10. Lower bound of u_T (r_{ZT} = variable 9)	.44	.46	.39	.24
11. Upper bound of u_T ($r_{ZT} = 0$)	.53	.59	.52	.45

*This person is assumed to be randomly selected from the DOL sample.

**The minimum value of r_{ZT} for black females is not a real number. The value of r_{UZ} was used as a proxy for r_{ZT}.

NOTE: The more conservative estimate of u_T (variable 10) is used to calculate the bias.

TABLE C-II

ESTIMATES OF THE PARAMETERS USED TO CALCULATE ℓn_T - THE DIFFERENTIAL PROBABILITY OF LOSING JOB TENURE BETWEEN TRAINEES AND CONTROLS

		White male	White female	Black male	Black female
1.	Maximum correlation between ℓn and Z - ($r_{\ell n Z}$)	.22	.19	.23	.20
2.	Probability that an individual* is a trainee - (p_T)	.42	.42	.42	.42
3.	Probability that a trainee* loses job tenure in 1963 - ($p_{\ell n/T=1}$)	1.0	1.0	1.0	1.0
4.	Probability that a control* loses job tenure in 1963 - ($p_{\ell n/T=0}$)	.19	.13	.20	.14
5.	Probability that an individual loses job tenure - ($p_{\ell n}$)	.61	.55	.62	.56
6.	Standard deviation of the trainee probability - (S_t) (variable 2)	.49	.49	.49	.49
7.	Standard deviation of the job tenure probability (variable 5) - ($S_{\ell n}$)	.49	.50	.49	.50
8.	Correlation between ℓn and T - ($r_{\ell n T}$)	.68	.77	.67	.75
9.	Estimate** of ℓn_T ($r_{ZT} = r_{\ell n Z}$)	.66	.77	.64	.75

*This person is assumed to be randomly selected from the DOL sample.

**The values of r_{ZT} that minimize u_T are not within the possible bounds of a correlation coefficient. Variations in r_{ZT} over a realistic range is reasonably small. Therefore only a single value for ℓn_T is shown.

APPENDIX D

THE DATA USED IN THIS STUDY

The data used in this study were derived from a stratified random sample of the Continuous Work History Sample (CWHS) of the Social Security Administration for the years 1959-1963. The CWHS is a set of longitudinal records of the earnings of over 800,000 individuals, derived from quarterly employer reports for a 1% random sample of workers covered by social security insurance (OASDHI). The initial sample used here was drawn for a different study and, while it included 10% of CWHS-covered individuals aged 20-60, it intentionally over-represented age groups 17-19 and 61-65. A number of steps were taken to reduce this initial sample to the final form used in this study.

The first step in reducing the initial sample was to select only individual records containing employer reports for each year covered by the sample. (Many individuals in the sample were out of the labor force or in employment uncovered by OASDHI for one or more years.) After this first reduction, the remaining records were divided into two job mobility groups: individuals were considered mobile if either the industry or county of major job was changed at some time during the period; otherwise they were considered non-mobile.* The final step was to select only workers between the ages of 23 and 53 as of 1959, whose earnings did not exceed the social security taxable limit of $4800 in any year.

These criteria were all intended to eliminate records of individuals who might be very different from manpower program enrollees. These steps also maximized the information available for each individual. Since quarterly earnings are not reported after the yearly total reaches $4800, the inclusion of individuals with earnings greater than $4800 would have limited the applicability of the unemployment measure used in this study. Similarly, if individuals with no employer record were included in the sample, the accuracy of the mobility measure would have been greatly reduced. In addition, this step decreased the possibility of failing to count earnings from non-covered occupations.

*Individuals who held two jobs in the same industry and county during a single year were also counted as mobile, although they possibly held both jobs simultaneously.

Finally, the prime age criterion was set to eliminate from the sample, young persons who were attending school, and older individuals in partial or full retirement.

Table D-I presents information about the sample size of the data used in this study.

The variables used in this study are defined in table D-II. In two cases the variables deserve further comment.

Quarterly Unemployment

The quarterly unemployment measure is an estimate of the amount of time an individual was not employed. It was constructed by detecting the dip in earnings that in most cases accompanies unemployment. Thus, a person is considered to have experienced unemployment in any quarter for which earnings were half the earnings of the highest quarter that year. In addition, a person was considered to have experienced unemployment in any quarter in which his earnings were below $300 because it was assumed that any individual earning less than $25 a week could not have been fully employed. The minimum earnings criterion was especially designed to provide a mechanism for eliminating part-time workers, particularly students, from the sample.

Loss of Job Tenure

The loss of job tenure is measured by detecting a change in major employer between two years rather than within a single year. (This latter measure could only be derived from quarterly data not readily available at this time.) In order to show a change of major employer between two years there must have been sufficient longevity of employment for one firm to have paid the largest proportion of earnings in the first year and a different firm to have paid the largest proportion of earnings in the second year. In order to satisfy this criterion it is most likely that employment with the first firm will have terminated roughly within the middle twelve months of the two-year period. Because other variables, particularly earnings, are measured within a single year, the longevity measures are closely correlated with variables in two different years.

In addition, it is not possible unequivocally to detect a job change without a change of industry or county of major employer, since this measure cannot differentiate between an individual who changes from one job to another while remaining in the same industry and county, and an individual who holds the same two jobs simultaneously.

Table D-III describes, in detail, the characteristics of the individuals in the sample.

TABLE D-I

SAMPLE SIZE BY SEX, RACE, AND MOBILITY GROUPS

		White Male	White Female	Black Male	Black Female
1.	Total individuals in the mobile group meeting age and income criteria	6385	5559	1773	851
2.	Percent of the total mobile group	45.2	38.2	12.2	5.8
3.	Total individuals in the non-mobile group meeting age and income criteria	2448	4832	715	753
4.	Percent of the total non-mobile group	28.3	55.0	8.2	8.6
5.	Combined total meeting criteria	8833	10390	2488	1604

		Mobile Group	Non-Mobile Group
6.	Total	53,849	31,985
7.	Total meeting age-income criteria	14,568	8,748
8.	Percent of total	27.0	27.4
9.	Total failed age criterion	23,794	7,981
10.	Percent of total	44.2	25.2
11.	Total failed income criterion but passed age criterion	15,486	15,256
12.	Percent of total	28.8	47.8

TABLE D-II

DESCRIPTION OF VARIABLES USED IN THIS STUDY

1. Years covered in the data: 1 = 1959, 2 = 1960, ..., 5 = 1963.

2. Age as of 1959: 3 = 23, 4 = 28, ..., 9 = 53
(There are no individuals aged 24-27, 29-32, etc. in the sample.)

3. Sex: 0 = male, 1 = female.

4. Race: 0 = white, 1 = non-white.

5. YUi - earnings in year i ($100's)

6. QUi - number of quarters unemployed in year i (earnings dip measure)

   A person is considered to experience a spell of unemployment in quarter j of year i if the earnings in quarter j are either less than half the earnings of the highest paid quarter or less than $300.

7. LNi - loss of job tenure dummy (job change dummy)

   A person is considered to have lost job tenure if he changed employers between year i and year i+1.

8. DQUi - quarterly earnings dummy, year i

   This variable indicates if an individual's earnings record satisfied any of the quarterly unemployment criteria listed for variable 6 (QUi). In that case DQUi = 1. Otherwise DQUi = 0.

TABLE D-III

AVERAGE CHARACTERISTICS OF THE SAMPLE BY RACE-SEX-MOBILITY GROUP

	White-male			White-female			Black-male			Black-female		
	Both groups	Mobile group	Non-mobile group	Both groups	Mobile group	Non-mobile group	Both groups	Mobile group	Non-mobile group	Both groups	Mobile group	Non-mobile group
Average age	35.3	33.5	40.0	38.95	37.5	40.6	35.0	33.7	38.4	37.5	36.2	39.1
Average earnings (dollars)												
Year 1959(Y1x100)	2331	2167	2759	2031	1018	2319	1988	1818	2410	1443	1247	1771
Year 1960(Y2x100)	2405	2208	2919	2246	2013	2514	2071	1873	2562	1626	1440	1836
Year 1961(Y3x100)	2495	2314	2967	2330	2071	2628	2120	1934	2581	1710	1555	1887
Year 1962(Y4x100)	2675	2504	3121	2470	2215	2763	2331	2164	2745	1813	1662	1984
Year 1963(Y5x100)	2780	2641	3142	2488	2257	2754	2436	2306	2758	1865	1743	2003
Average quarters–unemployed												
Year 1959(QU1)	1.25	1.39	.88	1.18	1.48	.83	1.29	1.48	.82	1.79	2.07	1.47
Year 1960(QU2)	1.19	1.36	.75	.98	1.28	.63	1.21	1.45	.61	1.60	1.87	1.30
Year 1961(QU3)	1.11	1.26	.72	.96	1.24	.64	1.18	1.40	.63	1.52	1.74	1.27
Year 1962(QU4)	1.06	1.20	.69	.89	1.14	.60	1.06	1.24	.61	1.45	1.62	1.26
Percent experiencing income dip greater than 25 percent												
Year 1959(YU1)	10.9	23.2	7.4	12.8	17.8	6.7	18.3	23.1	6.3	14.0	18.8	8.7
Year 1960(YU2)	16.6	20.5	6.4	15.1	21.0	9.4	18.1	22.4	7.4	15.6	20.5	10.9
Year 1961(YU3)	13.9	17.2	5.3	12.7	18.0	5.3	14.1	17.5	5.6	14.3	19.9	7.9
Percent unemployed												
Year 1961(DQU3)	46.3	53.0	28.8	41.3	51.6	28.4	48.9	57.5	27.5	54.5	63.4	44.6
Year 1962(DQU4)	44.6	50.9	28.2	39.5	50.0	26.4	44.3	52.1	24.9	53.6	61.5	44.7
Income distribution 1959												
$ 0-$1600	29.5	34.4	16.7	35.9	44.7	24.6	36.5	42.5	21.8	58.3	66.6	48.9
$1600-$3200	41.2	42.1	38.8	43.4	38.9	47.8	47.3	45.9	50.7	34.9	29.9	40.5
$3200-$4800	29.3	23.5	44.4	20.5	16.3	25.4	16.2	11.6	27.6	6.8	4.3	9.6
Percent changed job												
Year 1959(LN1)	36.7	50.7(2)	0	25.1	46.9(2)	0	37.0	53.2(2)	0	24.7	46.5(3)	0
Year 1960(LN2)	46.7	64.7(1)	0	25.8	48.1(1)	0	43.2	60.7(1)	0	25.7	48.5(1)	0
Year 1961(LN3)	33.1	46.0(3)	0	22.7	42.1(3)	0	35.2	49.5(3)	0	23.5	44.2(4)	0
Year 1962(LN4)	33.2	45.0(4)	0	22.3	41.9(4)	0	34.8	47.8(4)	0	25.9	48.9(2)	0
Distribution of last job change												
Between states												
Between industry(MB1)	.139	.197(2)	0	.056	.105(4)	0	.106	.149(3)	0	.036	.067(4)	0
Within industry(MB2)	.142	.194(3)	0	.023	.041(6)	0	.092	.129(5)	0	.022	.041(6)	0
Within states between counties												
Between industry(MB3)	.113	.154(4)	0	.077	.143(3)	0	.100	.142(4)	0	.077	.146(3)	0
Within industry(MB4)	.033	.046(6)	0	.042	.077(5)	0	.033	.047(6)	0	.028	.053(5)	0
Within counties												
Between industry(MB5)	.204	.284(1)	0	.227	.425(1)	0	.258	.362(1)	0	.178	.336(2)	0
Within industry(MB6)	.092	.126(5)	0	.110	.207(2)	0	.122	.172(2)	0	.189	.356(1)	0
Percent with three or more job changes (MB3)	24.6	34.1	0	11.5	21.2	0	25.8	36.1	0	13.7	25.7	0
Percent experiencing no job change 1959-1963	27.7			46.5			28.7			46.9		

Note: Numbers in parentheses indicate rank order.

D-5