NBER WORKING PAPER SERIES

THE PRE-PROGRAM EARNINGS DIP AND THE DETERMINANTS OF PARTICIPATION IN A SOCIAL PROGRAM: IMPLICATIONS FOR SIMPLE PROGRAM EVALUATION STRATEGIES

James J. Heckman
Jeffrey A. Smith

Working Paper 6983
http://www.nber.org/papers/w6983

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
February 1999

This research was supported by NSF SBR 91-11-455 and SBR 93-21-048 and by grants from the Russell Sage Foundation, the American Bar Foundation and the Social Science and Humanities Research Council of Canada. We thank Jingjing Hsee for her programming work, Theresa Devine for her comments and assistance with the SIPP data, Karen Conneely and Edward Vytlacil for their excellent research assistance and seminar participants at the October 1993 NBER Labor Studies Group meeting, Carnegie-Mellon University, the University of Western Ontario, Queen's University and McMaster University for their comments. The views expressed in this paper are those of the authors and do not reflect those of the National Bureau of Economic Research.

The Pre-Program Earnings Dip and the Determinants of Participation in a Social Program: Implications for Simple Program Evaluation Strategies
James J. Heckman and Jeffrey A. Smith
NBER Working Paper No. 6983
February 1999
JEL No. J24

ABSTRACT

The key to estimating the impact of a program is constructing the counterfactual outcome representing what would have happened in its absence. This problem becomes more complicated when agents self-select into the program rather than being exogenously assigned to it. This paper uses data from a major social experiment to identify what would have happened to the earnings of self-selected participants in a job training program had they not participated in it. We investigate the implications of these earnings patterns for the validity of widely-used before-after and difference-in-differences estimators. Motivated by the failure of these estimators to produce credible estimates, we investigate the determinants of program participation. We find that labor force status dynamics, rather than earnings or employment dynamics, drive the participation process. Our evidence suggests that training programs often function as a form of job search. Methods that control only for earnings dynamics, like the conventional difference-in-differences estimator, do not adequately capture the underlying differences between participants and non-participants. We use the estimated probabilities of participation in both matching estimators and a nonparametric, conditional version of the differences-in-differences estimator and produce large reductions in the selection bias in non-experimental estimates of the effect of training on earnings.

James J. Heckman
Department of Economics
University of Chicago
1126 East 59th Street
Chicago, IL 60637 USA
and NBER
jjh@uchicago.edu

Jeffrey A. Smith
Department of Economics
University of Western Ontario
Social Science Centre
London, Ontario N6A 5C2 Canada
and NBER
jsmith@julian.uwo.ca

The key to estimating the impact of a program is constructing the counterfactual outcomes representing what would have happened in its absence. This problem becomes more complicated when agents, such as individuals, firms or local governments, self-select into the program rather than being exogenously assigned to it. In many cases, agents self-select on the basis of the outcome variable that the program is designed to affect, as when trainees choose to take training when their earnings are low, or when states reform their social assistance systems in response to increases in the caseload. This can lead to selection bias in evaluating the program.

This paper examines a prototypical job training program into which participants self-select. It uses data from a major social experiment to identify what would have happened to the earnings of participants in a job training program had they not participated. We investigate the implications of these earnings patterns for the validity of widely-used before-after and difference-in-differences estimators. We demonstrate that these estimators do not produce credible estimates of the impacts of training. This leads us to investigate the determinants of program participation. We find that labor force dynamics, rather than earnings or employment dynamics, drive the participation process. For women, dynamic family processes related to marriage and childbearing are also important. Our evidence suggests that training programs function as a form of job search for many of their participants. Evaluation methods that only control for earnings dynamics, like the conventional difference-in-differences estimator, do not adequately capture the underlying choices leading to differences in unobserved variables between participants and non-participants. Application of our findings about the participation process in either matching estimators or a conditional (on the probability of participation) nonparametric version of the difference-in-differences estimator yields large reductions in the extent of selection bias in non-experimental estimates of the effect of training on earnings.

Historically, evaluators of early U.S. job training programs used before-after comparisons of participant earnings. The problem with this approach is that it attributes all improvements in outcomes relative to pre-program levels to the program being evaluated. If there are general increases in earnings due to economy-wide effects or life-cycle earnings growth, then this estimator will be biased.

To address this problem, it became common to utilize a comparison group of non-participants to eliminate common life-cycle and economy-wide factors from the before-after estimator. Such methods were widely used in the literature on evaluating educational interventions (see, *e.g.,* Campbell and Stanley, 1966). In the conventional difference-in-differences approach, the before-after earnings change for participants is compared to the before-after change for a temporally aligned group of non-participants. In the context of evaluating training programs, Ashenfelter (1978) noted a potentially serious limitation of this procedure when he observed that the mean earnings of participants in government training programs decline in the period prior to program entry. Subsequent research finds this regularity, sometimes called "Ashenfelter's dip" or the "pre-program dip", for participants in many other training and adult education programs (see Ashenfelter and Card, 1985, Bassi, 1983, 1984, and the comprehensive survey by Heckman, LaLonde and Smith, 1999).

Whether the pre-program drop in earnings is permanent or transitory determines what would have happened to participants had they not participated. Knowing whether the dip is permanent or transitory has important implications for the validity of both the before-after and conventional difference-in-differences estimation methods. Furthermore, the validity of variants of the conventional difference-in-differences approach that control for earnings histories depends on the relationship between earnings in the post-program period and the determinants of program participation.

Analysts of training programs using non-experimental data can only speculate about what the earnings of participants would have been had they not participated. In this paper, we use data on the control group from the National JTPA Study (NJS), a recent experimental evaluation of a large scale U. S. training program to learn what the earnings of participants would have been had they not participated.[1] The Job Training Partnership Act (JTPA) program is typical of many government job training programs around the world in terms of both its target population and the types of services it provides (see Heckman, LaLonde and Smith, 1999). Control group members were eligible for, applied to and were initially accepted into the JTPA program prior to being randomized out. Under certain conditions, their earnings represent the desired counterfactual.[2] For adult males, the control group data reveal that the dip in mean earnings is primarily

[1] See Bloom, *et al.* (1993) and Bloom, *et al.* (1997) for descriptions of the National JTPA Study.

[2] Heckman (1992), Heckman and Smith (1993, 1995) and Heckman, LaLonde and Smith (1999) discuss in detail the conditions under which experimental control group data provide the desired counterfactual. In short, these conditions are: (1) that random assignment be correctly conducted, so that control group members do not receive the experimental treatment; (2) that there is no "randomization bias" such that the program operates differently or serves different persons due to random assignment; and (3) that the control group members not receive subsitute treatments from other sources that are similar to the experimental treatment. Bloom (1991) and Bloom, *et al.* (1993) provide evidence in support of (1) for the NJS. Heckman, Khoo, Roselius and Smith (1996) report that there is little evidence of randomization bias in the JTPA experiment. Heckman, Hohmann, Smith and Khoo (1998) present evidence of violations of (3) for those control group members (about a third of the total) recommended to receive classroom training. Substitution is fairly low among the remaining controls.

transitory. For the other demographic groups considered, control group earnings grow above pre-program levels in the period following random assignment.

We show that this post-random-assignment earnings growth among the controls imparts a strong upward bias to before-after estimators of program impact. Early evaluators who used these estimators falsely attributed to the programs being evaluated improvements in earnings that would have occurred even in the absence of training.

The same bias plagues conventional difference-in-differences estimators. We apply these estimators to two comparison groups composed of persons eligible for JTPA. The first consists of eligible non-participants (ENPs) at four training centers in the JTPA experiment. In many ways, this comparison group is ideal. The ENPs reside in the same local labor markets as the experimental treatments and controls, complete the same surveys, and are all eligible for JTPA.[3] The second comparison group of eligibles is drawn from the 1986 Full Panel of the Survey of Income and Program Participation (SIPP). This sample resembles those used in earlier evaluations with the exception that program eligibility can be precisely determined in the SIPP data because there is much more information on monthly income dynamics.[4]

Compared to the experimental impact estimates, the conventional difference-in-differences estimators applied to either comparison group produce substantially biased estimates of program impacts because the upward trend in post-program earnings for controls is not found for comparison group members. That the earn-

[3]Heckman and Roselius (1994a) show that most comparison groups used in practice lack at least one, and often all, of these features. Heckman, Ichimura, Smith and Todd (1998) present evidence on the importance of using the same survey instruments and drawing participants and comparison group members from the same local labor market.

[4]See the Appendix for a more detailed data description.

ings behavior of the comparison groups does not correspond to that of the controls indicates that these groups do not provide the desired counterfactual. Furthermore, the earnings growth among controls after random assignment, along with the pre-program dip, makes the difference-in-differences estimator quite sensitive to the specific periods over which "before" and "after" are defined.

The failure of simple comparison group estimators suggests that the design of successful estimators may benefit from a deeper understanding of the program participation process. Partly due to data limitations, early analysts focused on earnings as the outcome measure of interest and on declines in the opportunity cost of taking training as the key determinant of program participation (see Heckman, 1978, Ashenfelter and Card, 1985, and the survey in Heckman, LaLonde and Smith, 1999). Even if this model is a valid description of the program participation process, conditioning on eligibility does not suffice to make comparison group members comparable to controls. While eligible adults sometimes experience a dip in earnings prior to the decision to participate in the program, their dip differs from that experienced by the controls in both its timing and intensity. The two dips differ because the dip among the controls results primarily from unemployment dynamics while the dip among the eligibles results primarily from reductions in earnings conditional on employment (Smith, 1997a). This mismatch helps account for the bias and instability in the conventional difference-in-differences estimator applied to earnings gains.

Unemployment dynamics and not earnings or employment dynamics, drive participation in training programs. Unemployment dynamics are only weakly related to earnings dynamics. For example, persons who re-enter the labor force and become unemployed have no change in their earnings but increase their likelihood of participation in training programs. Job training programs such as JTPA appear to operate as a form of job search. This is not surprising given that many of the services they offer – such as job search assistance and on-the-job training at private firms – are designed to lead to immediate employment.

We also show that a number of additional factors such as age, schooling, marital status and family income are important determinants of program participation. Our evidence explains the failure of econometric methods based on the assumption that earnings histories drive program participation and suggests the value of investigating alternative econometric strategies that exploit information on unemployment dynamics along with the additional factors determining program participation to control for self-selection bias in program participation.

We use our model of program participation to develop cross-sectional matching estimators and a conditional (on the probability of participation) nonparametric version of the difference-in-differences estimator that improve on the performance of conventional before-after and difference-in-differences methods. These methods reduce the estimated selection bias compared to what is obtained from conventional methods. Conditioning on labor force status histories plays a crucial role in reducing selection bias for adult men. However, substantial bias still remains.

The plan of the paper is as follows. Section 2 presents new evidence on the earnings of the experimental control group from the National JTPA Study. Section 3 compares the earnings dynamics of the comparison group samples to those of the controls and indicates the implications of their differing earnings patterns for the design and performance of difference-in-differences estimators of program impact. Section 4 analyzes the determinants of program participation. Section 5 demonstrates that the richer models of program participation are effective in reducing selection bias in non-experimental estimates of program impact. The final section summarizes the implications of our analysis for future evaluations of labor market programs.

2. The Pre-program Dip and the Before-After Estimator

In this section, we examine the mean earnings of randomized-out JTPA participants, and consider their implications for before-after estimators of program impact. Figures 1A to 1D display the mean earnings of eligible applicants accepted into the program but randomly denied access to services. This group is labelled "Controls" in Figures 1A to 1D.[5] Month 't' in this case represents the month of random assignment, which coincides with the month of eligibility determination for most controls.[6] The data show a large dip in the mean earnings of control group members for all four demographic groups: adult males and females (age 22 and above) and male and female out-of-school youth (ages 16-21). In each case, the dip reaches its lowest point in the month of random assignment. Ashenfelter (1978) first noted this pre-program dip in the earnings of participants in job training programs, and it has subsequently been found to be a feature of virtually all training and adult education programs (Heckman, LaLonde and Smith, 1999).

The pattern of recovery from the pre-program dip has important practical consequences for the performance of before-after estimators, which compare post-program earnings to pre-program earnings to measure the effect of the program. Define the impact of the program as the effect of the program on participants, compared to what they would have earned without participating in the program.

[5] Patterns are similar for the full set of 16 training centers in the National JTPA Study. We focus on the four centers at which the ENP sample was drawn.

[6] In some cases, lags in the intake process may cause the month of random assignment to lie one or two months after the month of eligibility determination.

If the decline in earnings prior to month 't' is transient, before-after comparisons will *overstate* the impact of the program on earnings if the earnings decline occurs in the period used to measure pre-training earnings. On the other hand, if the decline in mean earnings is persistent, before-after comparisons will *understate* the impact of the program if the decline occurs during the period used to measure pre-training earnings. Figure 1A reveals that the mean earnings decline for adult males is largely transient, while Figures 1B to 1D reveal that post-program earnings grow well above pre-program levels for the other three demographic groups. The timing of the earnings dip indicates that valid before-after comparisons will require more than a year of pre-program data for adults. Even with sufficient pre-program data, post-program earnings growth among controls implies a large upward bias in before-after estimators of program impact for all groups but adult males.

More precisely, let Y_{0a} denote earnings without training in the period after month 't' ($a > t$) and Y_{1a} denote earnings with training in the period after month 't'. Let $D = 1$ for persons who apply and are accepted into JTPA and $D = 0$ otherwise, and let $R = 1$ for persons who are randomized into the experimental treatment group (conditional on $D = 1$) and $R = 0$ for persons randomized into the experimental control group. Then the experimental impact is defined as:

$$E(Y_{1a} \mid R = 1, D = 1) - E(Y_{0a} \mid R = 0, D = 1). \tag{1}$$

This is just the difference in mean earnings between the experimental treatment and control groups in period a after random assignment. Under the assumptions that justify random assignment, this parameter estimates the effect of treatment on the treated.[7,8]

[7] See Heckman (1992), Heckman and Smith (1993, 1995) and Heckman, LaLonde and Smith (1999) for discussions of these assumptions.

[8] Heckman (1997), Heckman and Smith (1998) and Heckman, LaLonde and Smith (1999)

The non-experimental before-after estimator converges to:

$$E(Y_{1a} \mid R = 1, D = 1) - E(Y_{0b} \mid R = 0, D = 1), \tag{2}$$

where the subscript b denotes the period before month 't'$(b < t)$, and where the pre-random-assignment earnings of the control group are used to proxy the pre-random-assignment earnings of the treatment group. The mean bias that results from using (2) in place of (1) to estimate the impact of treatment on the treated is

$$E(Y_{0a} \mid R = 0, D = 1) - E(Y_{0b} \mid R = 0, D = 1), \tag{3}$$

the difference between the pre-program earnings of participants and what their post-program earnings would have been, had they not participated.

Table 1 presents before-after estimates based on the JTPA data. The first two columns define the "before" and "after" periods used in each estimator. The experimental impact for the "after" period is given in column three and the before-after estimates appear in column four. (The numbers in the remaining columns are defined later in this paper). For all demographic groups and for all base periods, the before-after estimate substantially exceeds the experimental estimate. For example, the eighteen-month before-after estimate for adult males is $3,109 compared to an experimental estimate of $657. Note that, for adult males, the group with the largest pre-program dip and the smallest (relative) post-program earnings growth for the controls, the before-after impact estimate is substantially larger for before periods that include the pre-program dip than for those that do not.

---

consider the limitations of this parameter and discuss other parameters of interest to evaluators.

Can we do better by using a difference-in-differences estimator? We now address this question using two comparison groups selected according to the intuitively appealing criterion that included persons be eligible for the program but not participate in it.

3. Comparison Groups and the Conventional Difference-in-Differences Estimator

In this section, we first describe the eligibility rules for JTPA that define our two non-experimental comparison groups of eligibles. We show that the earnings patterns of the two comparison groups differ in important ways from the pattern found for controls. Conditioning on eligibility alone does not eliminate bias. As a result, conventional difference-in-differences estimators based on comparison groups of eligibles are both biased and unstable.

A. The JTPA Eligibility Rules

Economic disadvantage is the primary eligibility condition for JTPA training. It consists of an individual having either low family income in the six months prior to application or current participation in a means-tested social program.[9,10]

[9] As defined in the Job Training Partnership Act, economic disadvantage arises if at least one of the following criteria are met: (1) low *family* income in the six months prior to application; (2) current receipt of cash public assistance such as Aid to Families with Dependent Children (AFDC) or general assistance; and (3) current receipt of food stamps. According to the U.S. Department of Labor (1993a), in Program Year 1991 (July 1991 to June 1992), around 93 percent of JTPA participants qualified because they were economically disadvantaged. Similar measures of economic disadvantage have formed the basis of eligibility for most U.S. job training programs.

[10] A second, and much less important, avenue to JTPA eligibility is an "audit window" that allows up to 10 percent of participants at each training center to be non-economically-disadvantaged persons with other barriers to employment such as limited ability in English. Due to the subjective nature of these barriers, the eligibles examined here consist only of persons eligible via economic disadvantage.

The key features of the eligibility rules are the dependence on family (rather than individual) income and the short six-month window over which income is summed to determine eligibility. The six-month window allows highly-skilled workers to become eligible for the program after only a few months out of work.[11,12]

Barnow (1993) shows that there are slight differences between the eligibility criteria for JTPA and those of its predecessor programs.[13] All major training programs in the U.S. have focused on displaced workers, persons with low incomes and transfer recipients. Furthermore, our evidence on the determinants of participation suggests that many differences in eligibility rules across programs will have little impact on the types of persons participating. In particular, because recently unemployed persons and persons re-entering the labor force are much more likely to select into the JTPA program than other eligibles, differences in eligibility rules across programs that do not affect the eligibility status of these groups will have only a limited effect on the composition of program participants.

B. Comparing the Earnings Patterns of Eligibles and Participants

We now compare the mean earnings patterns of the two comparison groups of program eligibles and the experimental controls from the National JTPA Study.

[11] Devine and Heckman (1996) present an extensive discussion of the JTPA eligibility rules, their variation over time and across states, and the implications of this variation for the composition of the eligible population.

[12] The implementation of the general rules described here varies somewhat across localities, as states and training centers have some discretion over exactly what constitutes family income and what constitutes a family. Devine and Heckman (1996) show that such differences are too small to affect the patterns discussed here. The eligibility rules described here are those in place at the time our data were collected. Since that time some marginal changes have been made. See Devine and Heckman (1996) or U.S. Department of Labor (1993b).

[13] The JTPA program replaced the CETA (Comprehensive Employment and Training Act) program which had earlier replaced the MDTA (Manpower Development and Training Act) program.

The eligible non-participant, or "ENP", comparison group is drawn from the same local labor markets as the controls and has earnings data collected using the same survey instrument. In contrast, the "SIPP" comparison group is a national sample drawn from a major U.S. panel data set. Both comparison groups are composed exclusively of persons eligible for JTPA. Differences in the time series of mean earnings between a comparison group and the controls generally produce bias in the difference-in-differences estimator.

Figures 1A to 1D display the mean individual earnings of the controls and the two comparison groups of eligibles. For the comparison groups, month 't' is the month of measured eligibility. Adult male and adult female SIPP eligibles display a dip in mean earnings centered in the middle of the six-month window over which components of family income earnings are summed to determine JTPA eligibility, although the dip for women is much less pronounced. Devine and Heckman (1996) prove that the JTPA eligibility rules generate such a dip for stationary family income processes; since adult earnings are typically a large component of family income in low-income families, this pattern also shows up in graphs of individual earnings for adult eligibles. In contrast, youth in the SIPP eligible sample experience no dip in mean individual earnings. These demographic differences in earnings dynamics indicate that, except for adult males, eligibility depends crucially on the earnings behavior of other family members. For adult SIPP eligibles, mean earnings recover from their decline because the eligibility rules for JTPA (and many other programs) operate to include persons temporarily suffering adverse economic circumstances.

Comparing the mean earnings of the SIPP eligibles to those of the JTPA controls from the National JTPA Study, we find substantial differences between the two groups. Among adults, the magnitude of the dip is larger for the controls, whose dip is centered at month 't' rather than three or four months earlier. Among youth, only the controls show any dip at all. This evidence strongly suggests that while the JTPA eligibility rules clearly affect the mean earnings patterns observed for all eligibles, additional behavioral factors are required to account for the dip observed for program participants.

Adult male and female ENPs show no dip in mean earnings during the period prior to month 't.' Smith (1997a) demonstrates that the absence of a dip for this group results from the structure of the survey instrument used to gather earnings information on the ENPs. This survey instrument smooths away all within-job variation in earnings. Such variation is an important component of the dip observed among the SIPP eligibles. It plays a relatively small role in the earnings dip for the controls, most of which results from the effects of job loss, which are captured by the survey. A better survey would have revealed a greater decline in earnings for both the ENPs and the controls.[14] Furthermore, with the exception of male youth, the ENPs do not experience earnings growth after month 't' to match that found for the controls or, to a lesser extent, for the SIPP eligibles.

The differences between the earnings patterns of the controls and the two comparison groups prior to month 't', and the post-'t' divergence in mean earn-

[14] See Smith (1997a,b) and the extended data appendix (available on request) for a more detailed discussion of these issues and of the difference in mean earnings levels between the SIPP eligibles and the ENPs. In brief, differences in observed characteristics do not explain the difference in mean earnings levels between the ENP and SIPP samples of eligibles. Instead, non-response bias among the ENPs (low income persons are more likely to attrit from the sample), local labor market factors, differences in the distribution of calendar months of eligibility and differences in the way the underlying survey instruments measure hours worked and income from overtime, tips, bonuses and commissions account for the differences in mean earnings levels, with the relative importance of these factors varying by demographic group.

ings due to earnings growth among controls, produce the failure of conventional difference-in-differences estimators. Using the notation already defined, the population version of the conventional difference-in-differences estimator is defined as:

$$[E(Y_{1a} \mid D=1, R=1) - E(Y_{0b} \mid D=1, R=0)] - [E(Y_{0a} \mid D=0) - E(Y_{0b} \mid D=0)]. \tag{4}$$

The estimator is implemented by replacing population expected values with their sample analogs.[15]

The last two columns of Table 1 present conventional difference-in-differences estimates of the impact of training on earnings constructed using the ENP (column 5) and SIPP eligible (column 6) comparison samples. These estimates reveal a general pattern of upward bias relative to the experimental impact estimates. Furthermore, the differences in the earnings patterns across groups – in particular the pre-program dip and post-random-assignment earnings growth experienced by the controls – lead to a high degree of sensitivity to the choice of "before" and "after" time periods used to generate estimates. For example, for adult males, the estimates using the twelve months before and after month 't' are dominated by the pre-program dip and so are positive. In contrast, the estimates using months 16 to 18 before and after month 't' are dominated by the post-random-assignment earnings growth of the controls, and so are negative.

Heckman, Ichimura, Smith and Todd (1998), Heckman and Todd (1996) and

[15] This estimator is widely used and a number of economists in the past decade have claimed credit for inventing it. Heckman and Robb (1985, 1986) discuss it among the many estimators they examine. Ashenfelter (1978) uses it and Campbell and Stanley (1966) discuss and apply it. See their Section 14 on multiple time series.

Heckman and Roselius (1994a,b) show that the failure of the conventional difference-in-differences estimator for these comparison groups persists when the estimates are adjusted for differences in observable characteristics.

4. The Determinants of Participation in JTPA

Heckman (1978) developed a model of program participation that is applied by Ashenfelter and Card (1985). The model is summarized in Heckman, LaLonde and Smith (1999). It focuses on earnings changes as determinants of participation. This emphasis was a natural consequence of Ashenfelter's discovery and reflects the limited data available to early analysts. This line of thought produced a set of longitudinal estimators that used earnings histories to eliminate differences between participants and non-participants. These estimators were extensively developed in Heckman (1978) and Heckman and Robb (1985). We have just shown that simple versions of these estimators are not effective. The evidence presented in Heckman, LaLonde and Smith (1999) indicates that more sophisticated versions do not perform any better.

A central principle of the evaluation literature introduced in Heckman and Robb (1985) is that knowledge of the determinants of program participation should guide the appropriate choice of a non-experimental estimator. The early literature assumed that earnings dynamics drove the participation process and used longitudinal estimators tailored to that assumption. A major finding reported in this paper is that it is unemployment dynamics that drive program participation and not earnings dynamics. Once this is recognized, progress can be made toward solving the problem of devising a good non-experimental estimator for evaluating job training and adult education programs.

A. The Important Role of Labor Force Status Dynamics

We now show that unemployment histories do a better job of predicting participation among eligibles than alternative measures based on earnings or employment, particularly for groups other than adult men. This evidence helps to account for the disappointing performance of econometric evaluation models that assume that program participation depends solely on earnings or employment histories.

The top panel of Table 2 presents participation rates calculated using the ENP and control data. Labor force status – whether a person is employed, unemployed or out of the labor force – plays a key role in determining the probability of participation in the JTPA program for all four demographic groups. In every case, those unemployed in the month of measured eligibility have by far the highest probability of application to, and acceptance into, the JTPA program.[16] This over-representation of the unemployed among participants implicitly suggests that participants place a fairly low value on the services provided by JTPA, as they are willing to participate, in general, only when the opportunity costs are low, because they are not working, and the benefits are high, because they are looking for work.

Going back over spells, we find that both the labor force status in the month of measured eligibility or random assignment and the labor force status in the preceding spell affect the probability of participation in JTPA. The two most recent labor force statuses during the period including month 't' and the six preceding months define a set of nine labor force status patterns. For example, the pattern labelled "Emp -> Unm" refers to persons who were unemployed in month

[16] Sandell and Rupp (1988) find that the unemployed have a higher probability of JTPA participation than the employed or those out of the labor force in their comparison of national samples of JTPA participants (drawn from administrative data) and program eligibles (constructed using the CPS).

't' but whose most recent labor force status during the preceding six months was employment. Repeated patterns such as "OLF -> OLF" indicate persons with the same labor force status in month 't' and in all six preceding months.

The bottom panel of Table 2 displays participation rates conditional on these labor force status transitions. Several interesting patterns emerge. First, substantial variation in participation rates exists among persons who do not work in any of the seven months up to and including month 't'. For all four demographic groups, the participation rate of persons persistently out of the labor force during this period lies well below that for persons unemployed for all seven months, and for persons who transit into or out of the labor force.

Second, for groups other than male youth, job leavers have a higher probability of program participation if they remain in the labor force after leaving their jobs than if they do not. Third, for adult females and male youth, program participation rates are higher among job gainers who found a job while unemployed than among those who found a job while out of the labor force. Finally, for adult males the participation rate of persons persistently out of the labor force substantially exceeds that of continuously employed persons. For the other three demographic groups, these two participation rates are roughly equal.

The importance of unemployment, and transitions into unemployment, as predictors of participation in JTPA is shown graphically in Figures 2A to 2D, which show the fraction of the ENPs and controls in each of the three labor force statuses – employed, unemployed and out of the labor force – in the months surrounding random assignment (for the controls) and measured eligibility (for the ENPs). For each of the four demographic groups, the fraction unemployed in the control group increases during the period leading up to month 't', as individuals transit into this status, with the result that the unemployed are over-represented among participants in each case.

B. Alternative Labor Market Variables

The National JTPA Study data contain a far richer set of variables than those available to previous analysts. Table 3 contrasts the data available from the NJS with that available in Ashenfelter (1978), Ashenfelter and Card (1985), LaLonde (1986), Bryant and Rupp (1987) and Dickinson, Johnson and West (1987). In this section we examine a variety of labor market variables to see which ones perform well using a common measure of predictive performance. We seek to determine the key behavioral determinants of participation, to form the cornerstone of an econometric model that successfully corrects for selection bias.

Table 5 summarizes our evidence on the performance of various predictors of program participation. Definitions of the variables used in the estimation appear in Table 4. We consider fifteen specifications broken down into four groups. The first group contains two specifications limited to background variables; these specifications serve as a benchmark. The remaining groups are for specifications based on employment, earnings and labor force status variables, respectively.

Each row of Table 5 presents the fraction of the control and ENP observations predicted correctly using a given set of regressors. Estimated standard errors for the prediction rate appear in parentheses. For each specification, separate equations are estimated and reported for each of the four demographic groups.[17]

[17] In an appendix available on request, we show that (1) the relative performance of the alternative specifications is robust to removal of the background variables; (2) the 0.03 cutoff value typically lies close to that which maximizes the equal-weights prediction rate; and (3) changing the cutoff value from 0.03 to either 0.01 or 0.05 does not affect the relative performance of the various labor market variables at predicting program participation.

The reported fraction of correct predictions consists of the simple average of the control and ENP correct prediction rates. This weighting is consistent with a symmetric loss function for misclassifications in the two groups.[18] A person is predicted to be a control if his or her estimated probability of participation exceeds 0.03, the assumed fraction of participants in the population.[19]

The first group presented in Table 5 includes specifications based solely on demographic background variables as defined in Table 4. The two specifications differ only in that the second includes a categorical family income variable. The first specification predicts remarkably well, especially for adult males. Adding family income improves the prediction rate for males but not for females.

The specifications in the second group include employment-related variables. The first specification in this group includes only an indicator variable for whether or not the person is employed in month 't'; this specification has a surprisingly high prediction rate. The second specification adds the background variables, which increases the prediction rate for all four groups relative to either the background variables alone or the employment indicator alone. The greater predictive power of the employment variables for adult males compared to the other three groups is a major finding, and motivates our search for other determinants of participation.

The third specification includes an employment transition variable that is similar to the labor force status transitions but combines persons who are unemployed

[18] Note that if the population-weighted prediction rate is used, then a correct prediction rate of 0.97 can be achieved by predicting everyone to be an ENP.

[19] Hunt, *et al.* (1984) estimate that 1.85 percent of persons eligible at some time during calendar year 1983 participated in JTPA. Sandell and Ruup (1988), using administrative data on JTPA participants from the Job Training Quarterly Survey for Program Years 1984 and 1985, along with data on persons eligible for JTPA constructed using the March 1986 CPS, estimate an annual participation rate of 2.3 percent among persons eligible at some time during a given year. This estimate may be broken down into separate estimates of 1.6 percent for adults age 22 to 64 and 5.1 percent for youth age 16 to 21.

and out of the labor force. For adult males, but not the other three groups, employment transitions do almost as well as labor force status transitions at predicting program participation. The fourth specification in this group, denoted 'ETC' in Table 5, includes a categorical variable based on the number of transitions from employment to non-employment (or vice-versa) in the twenty-four months prior to month 't'. This specification has the highest prediction rate overall for male youth. The last two specifications include categorical variables based on the number of job spells during the eighteen or forty-eight months prior to month 't'; these specifications perform relatively poorly.

The third group includes specifications based on earnings-related variables. The two specifications include monthly earnings in each of the six months prior to month 't' and quarterly earnings in each of the four quarters prior to month 't'. The earnings history variables predict program participation moderately well for adult males, but much less well for the other three groups. We examine a number of other earnings-based variables, including more complicated variables based on earnings patterns in the months prior to month 't', and none perform particularly well. For this reason, we do not discuss them here. Earnings patterns alone are relatively poor predictors of program participation, especially for groups other than adult males. This is not surprising for youth or re-entrant women who have no earnings. It is an important finding which helps account for the disappointing results reported in Ashenfelter and Card (1985), who implement a longitudinal non-experimental evaluation strategy using earnings histories.

Specifications including detailed labor force status variables comprise the final group in Table 5. The first two specifications include indicators for labor force status in month 't', with and without the background variables. Both predict far better than the corresponding specifications involving employment at 't.' Distinguishing between non-employed persons who are and are not looking for work (*i.e.*, between the unemployed and those out of the labor force) is crucial in successfully predicting program participation. The other three specifications incorporate variables measuring the dynamics of labor force status. For adult males, the specification based on the two most recent labor force statuses has the highest overall prediction rate. For women, it is a close second to the specification based on the amount of time in the most recent labor force status. For male youth, it is a close second to the specification based on the employment transition variable.

In comparing across specifications in Table 5, the prediction rates of certain pairs of specifications often cannot be statistically distinguished. However, the broad pattern of the table is clear. With the exception of the employment transition variable for male youth, specifications based on recent labor force dynamics that explicitly separate the unemployed from those out of the labor force do better at predicting program participation than those based on employment or earnings.

C. Multivariate Analysis

This section presents a multivariate analysis of the determinants of participation in the JTPA program conditional on eligibility using the data on experimental controls and ENPs from the National JTPA Study.[20] Our multivariate analysis reveals the central role of recent labor force status dynamics in determining program participation, as well as the contributing role of other factors such as age, schooling, marital status and family income. We focus on the labor force status

[20] Heckman and Smith (1997) present a more detailed analysis of the JTPA participation process in which the participation process is decomposed into a series of stages such as eligibility, awareness, application and acceptance into the program.

transition variables that we find do a better job of predicting participation in JTPA than other measures based on earnings or employment histories.

Table 6 reports estimates of logit models of participation in JTPA. The table includes coefficient estimates, estimated standard errors and mean numerical derivatives (or finite differences in the case of indicator variables). Coefficient estimates and estimated standard errors account for the choice-based nature of the sample.[21]

The results for adult males and adult females show that the coefficients for all eight of the labor force status pattern indicators are statistically significantly different from zero. For both groups, the smallest coefficient is on the indicator variable for those persistently out of the labor force; their participation probabilities differ the least from those of the persistently employed, who constitute the omitted group. The relative effects of the labor force status patterns parallel the ordering of univariate participation rates in Table 2.

Older adults have a lower conditional probability of participation which is consistent with the view that returns to training decline with age. The effect of completed schooling on the probability of participation shows a hill-shaped pattern, with adults with fewer than 10 or more than 15 years of schooling having differentially low estimated participation probabilities. Heckman and Smith (1997) show that this pattern results from low rates of program awareness among those with little schooling, and low rates of participation conditional on awareness among the highly educated.

Currently married adults of both sexes are relatively less likely to participate than those who have never married, while those whose marriage ended more than

[21] We use standard methods as exposited in Amemiya (1985).

two years ago are relatively more likely to participate. The effect is especially strong for adult women for whom training programs often provide a bridge back into the labor force following divorce. Receipt of food stamps has a positive effect on the participation probability for both groups, while participation in welfare (AFDC receipt) has a negative effect. Because nearly all AFDC recipients also receive food stamps, the coefficients on the AFDC receipt indicators should be interpreted as the effect of receiving both types of assistance rather than just food stamps. Finally, adult male eligibles with family incomes over $15,000 in the past year are relatively less likely to participate while adult females with family incomes between $3000 and $9000 have the highest probability of participation.

The same basic patterns are found for male and female youth. In particular, the labor force status pattern variables play a key role in determining participation for both groups.[22]

The fundamental importance of labor force status dynamics in determining participation is clearly evident even in a more general statistical model. For adult women, changes in the life cycle dynamics of the family, especially divorce, childbearing, and the entry of children into school are also important. A number of other factors including age, schooling, marital status and family income also help to determine participation for all demographic groups.

[22] Absent from the specifications reported here are measures of the state of the local economy at each of the four training centers. We estimated models including both county-level monthly unemployment rates averaged over the counties served by each center, and interactions between these unemployment rates and the center indicators. These variables never attained statistical significance and never had a noticeable impact on the proportion of correct predictions. One reason for this is that the number of ENPs whose month of measured eligibility occurs in a given calendar month depends not only on the size of the eligible population in that month, but also on the administrative schedule of the firm conducting the surveys. A second reason is that the flow into the program, as measured by the number of persons randomly assigned in each calendar month, depends on other factors beyond the local economy, including the academic schedule of the community colleges that provide much of the JTPA training at these centers.

5. *Selection Bias and the Determinants of Participation in JTPA*

In this section we show how the knowledge gained from our analysis of the determinants of participation in JTPA can be used to improve the performance of non-experimental evaluation methods in estimating the impact of JTPA on earnings. We focus here on two strategies that compare participants and non-participants based on their observed characteristics. The counterfactual for a given participant is estimated by the outcomes experienced by non-participants with the same or "similar" observable characteristics.

Both methods are based on "selection on observables" (see, *e.g.*, Heckman and Robb, 1985) and rest on assumptions regarding the relationship between earnings and program participation conditional on observed characteristics. They allow us to exploit in a structured way what we have learned about the determinants of participation in the preceding sections. The cross-section matching estimator assumes that conditional on a vector of observed characteristics X, D is independent of the non-participation outcome Y_{0a}. In formal terms, it is assumed that

$$(Y_{0a} \perp\!\!\!\perp D) | X,$$

where $\perp\!\!\!\perp$ denotes independence. As noted by Heckman and Robb (1986), and Heckman, Ichimura, Smith and Todd (1998), in order to use matching to estimate the parameter "treatment on the treated" it is only necessary to assume conditional mean independence so that conditional on X,

$$E(Y_{0a} \mid X, D = 1) = E(Y_{0a} \mid X, D = 0).$$

Selective differences in non-participation outcomes are assumed to be eliminated by conditioning on X. The nonparametric conditional difference-in-differences estimator introduced in Heckman, Ichimura and Todd (1997, 1998) and Heckman, Ichimura, Smith and Todd (1998) assumes that, conditional on X, selection bias in Y_0 is the same in particular periods before and after the participation decision, so that, conditional on X, it can be differenced out. In formal terms, the method assumes that in the population,

$$[E(Y_{0a} \mid X, D = 1) - E(Y_{0a} \mid X, D = 0)] -$$
$$[E(Y_{0b} \mid X, D = 1) - E(Y_{0b} \mid X, D = 0)] = 0,$$

so that (4) identifies $E(Y_{1a} - Y_{0a} \mid X, D = 1)$, the effect of treatment on the treated, where the a and b subscripts again denote periods before and after month 't.' As noted in Heckman and Robb (1985, 1986), this estimator assumes that common time (or age) effects operate on treatment and comparison group members so they can be differenced out.

Whereas matching is assumed to eliminate the bias in the post-program period, the conditional difference-in-differences estimator assumes the same cross-section bias in periods a and b so that differencing the outcomes between a and b eliminates the common bias component. Note that the conventional difference-in-differences estimator considered earlier is a crude version of this estimator in which the only conditioning variable is eligibility for JTPA.

Rosenbaum and Rubin (1983) demonstrate that under general conditions, conditioning on X is equivalent to conditioning on the probability of participation $\Pr(D = 1|X) = P(X)$. In this case, $P(X)$ replaces X in the assumptions justifying the matching estimator. Heckman, Ichimura and Todd (1997, 1998) develop a nonparametric difference-in-differences estimator that also conditions on $P(X)$. They develop the statistical properties of the matching and nonparametric difference-in-differences estimator when $P(X)$ is estimated.[23] We use their asymptotic theory to produce the estimates and standard errors reported below.

A general definition of the matching estimator for the impact of treatment on the treated presented in Heckman, LaLonde and Smith (1999) is:

$$\widehat{M_a}(S) = \sum_{i \in I_1} [Y_{1ai} - \sum_{j \in I_0} W_{N_0,N_1}(i,j) Y_{0aj}], \quad \text{for } P(X) \in S,$$

where Y_{1ai} denotes earnings with training in the post-program period for participant i, Y_{0aj} denotes earnings without training in the post-program period for non-participant j, N_1 is the number of program participants, N_0 is the number of persons in the comparison group, and I_1 and I_0 are sets of indices for participants and comparison group members, respectively. $W_{N_0,N_1}(i,j)$ is the weight attached to comparison group member j in constructing the counterfactual outcome for participant i. These weights sum to one for each participant so that $\sum_{j \in I_0} W_{N_0,N_1}(i,j) = 1$ for all i. The set S is the "common support" of $P(X)$ – that is, the subset of (0,1) for which values of $P(X)$ are present in both the participant and comparison group samples.[24] Matches for each participant are constructed by taking weighted averages over comparison group members.

Matching estimators differ in the weights they attach to members of the comparison group (Heckman, Ichimura and Todd, 1997). For example, "nearest-neighbor" matching sets all the weights equal to zero except for that on the comparison group observation with the estimated probability of participation closest

[23] Rosenbaum and Rubin (1983) assume that $P(X)$ is known rather than estimated.

[24] The region of common support consists of those values of $P(X)$ such that the smoothed densities of $P(X)$ in both the $(D = 1)$ and $(D = 0)$ samples are above a trimming level $\widehat{q}$. Formally, $S = \{P(X) : \widehat{f}(P(X)|D = 1) > \widehat{q} \text{ and } \widehat{f}(P(X)|D = 0) > \widehat{q}\}$, where $\widehat{f}$ is a smoothed density of $P(X)$ obtained using a standard kernel density estimator. Appendix C of Heckman, Ichimura and Todd (1997) discusses the choice of $\widehat{q}$ and reports that the bias estimates are sensitive to the value of $\widehat{q}$ only for small samples, such as that for male youth.

to that of the participant being matched, whose weight is set to one.

In contrast, the commonly-used kernel matching approach uses the weights:

$$W_{N_0,N_1}(i,j) = \frac{G_{ij}}{\sum\limits_{k \in I_o} G_{ik}},$$

where $G_{ik} = G((X_i - X_k)/a_{N_0})$ is a kernel function and a_{N_0} is a bandwidth parameter.[25] Kernel matching is a local averaging method that reuses and reweights all of the comparison group observations in constructing the estimated counterfactual outcome for each treatment sample member. Relative to nearest neighbor matching, kernel matching reduces the variance of the matching estimate by making use of information from additional non-participant observations. At the same time, it increases the bias in small samples because the additional observations are more distant, in terms of their probabilities of participation, from the observation being matched.

The matching estimates we report in Table 7 are based on the local linear matching method developed in Heckman, Ichimura and Todd (1997, 1998) and Heckman, Ichimura, Smith and Todd (1998). Local linear matching differs from kernel matching in the addition of a linear term in the probability of participation when constructing matches. To understand the method, note that one can construct the kernel estimate of the counterfactual outcome for participant i by running a weighted regression using all of the comparison group observations with non-zero weights with Y_{0aj} as the dependent variable. The regression contains only an intercept term and the estimated intercept is the kernel estimate of the counterfactual outcome for participant i. Local linear matching works the same

[25] a_{N_0} satisfies $\lim\limits_{N_0 \to \infty} a_{N_0} \to 0$. Precise conditions on the rate of convergence needed for consistency and asymptotic normality of the matching estimators used here are presented in Heckman, Ichimura and Todd (1997).

way except that the weighted regression for each participant i includes both an intercept term and a linear term in the probability of participation. This smooths out the estimate of the intercept and has desirable statistical properties if the underlying model is smooth.[26]

We use local linear weights instead of more conventional kernel weights because local linear estimators converge at a faster rate at boundary points and adapt better to different data densities. The boundary behavior is potentially important in our context because many observations in both groups have values of $P(X)$ close to the boundary value of zero.[27]

The conditional on $P(X)$ difference-in-differences estimator is defined as:

$$\begin{aligned}\widehat{D_{a,b}}(S) &= \sum_{i\in I_1}\left[(Y_{1ai}-Y_{0bi})-\sum_{j\in I_0}W_{N_0,N_1}(i,j)(Y_{0aj}-Y_{0bj})\right]\\ &= \widehat{M}_a(S)-\widehat{M}_b(S),\end{aligned}$$

where $\widehat{M}_b(S)$ is constructed using the same weights as $\widehat{M}_a(S)$ but is calculated using pre-program earnings data as the outcome measure.[28]

[26]The exact form of the weight for local linear matching is:

$$W_{N_0,N_1}(i,j) = \frac{G_{ij}\sum\limits_{k\in I_0}G_{ik}(P(X_k)-P(X_i))^2-[G_{ij}(P(X_j)-P(X_i))][\sum\limits_{k\in I_0}G_{ik}(P(X_k)-P(X_i))]}{\sum\limits_{j\in I_0}G_{ik}\sum\limits_{k\in I_0}G_{ij}(P(X_k)-P(X_i))^2-\left(\sum\limits_{k\in I_0}G_{ik}(P(X_k)-P(X_i))\right)^2}$$

[27]However, Heckman, Ichimura and Todd (1997) show that other matching methods yield qualitatively similar results.

[28]See Heckman, Ichimura and Todd (1997,1998) or Heckman, Ichimura, Smith and Todd (1998) for more detailed descriptions of both estimators and formal analyses of their statistical properties.

There are two ways in which what we learned in Section 4 about the determinants of program participation can help improve the performance of non-experimental evaluation methods. The first insight is that, for probabilities of participation based on the specification in Table 6, which incorporate the important labor force status transition variables, there are many participant ($D = 1$) observations for which the ENP comparison group contains no non-participant ($D = 0$) observations with the same or similar probabilities of participation. Put more simply, P-comparable non-participants are unavailable for many participants.

Heckman, Ichimura and Todd (1997) and Heckman, Ichimura, Smith and Todd (1996,1998) refer to this as the "common support" problem, as the empirical supports of the distributions of participation probabilities differ between participants and non-participants. They show that the failure of the common support condition accounts for a substantial fraction of the selection bias in simple cross-section comparisons of the earnings of participants and non-participants. Moreover, they show that imposing the common support condition substantially reduces selection bias in a variety of cross-sectional non-experimental evaluation procedures, including several variants of matching, conditional (on $P(X)$) difference-in-differences and the Heckman (1979) "two step" procedure.

The second way in which a better understanding of the determinants of participation helps to improve the performance of non-experimental evaluation methods is by providing probabilities of participation that are more likely to satisfy the assumptions underlying the matching and conditional difference-in-differences evaluation methods. The importance of understanding the determinants of participation is demonstrated by the results presented in Table 7.

Estimates

Table 7 presents estimates of the extent of selection bias (the difference between the experimental estimate and the estimate that would be yielded by a non-experimental estimation procedure) for the parameter treatment on the treated, $E(Y_{1a} - Y_{0a} \mid X, D = 1)$, when matching and conditional difference-in-differences methods are applied to the experimental controls and ENPs from the JTPA data to estimate the impact of JTPA participation on earnings in the 18 months after random assignment.[29,30] Table 7 contains four panels, one for each demographic group. For each demographic group, there are five columns and two rows. The first column presents the unadjusted mean difference in outcomes between the experimental controls ($D = 1$) and the ENPs ($D = 0$). This difference represents the selection bias present in the simple estimator that compares the unadjusted mean earnings of participants and eligible non-participants in the post-program period.

The remaining four columns present the selection bias present in non-experimental estimates based on probabilities of participation constructed using successively richer sets of conditioning variables X. The probabilities in the "Coarse I" col-

[29] We report the estimated selection bias here, rather than the impact estimates as in Table 1, because we lack the data on Xs required to construct the probabilities of participation for the experimental treatment group. This in turn means that we are unable to impose the common support condition on the treatment group data. The bias is calculated by replacing the earnings values of the experimental treatment group in the post-program period with the earnings values of the experimental control group in the same period in the formulas for the local linear matching and conditional difference-in-differences estimators.

[30] A similar analysis could be conducted using the SIPP as the source of the comparison group. See Heckman, Ichimura, Smith and Todd (1998) for estimates using this framework for three SIPP comparison groups constructed using different sample inclusion criteria. Heckman and Roselius (1994a,1994b) construct SIPP comparison samples based on matching the local labor market characteristics of the controls.

umn are based only on a small set of demographic variables.[31] The "Coarse II" column augments the demographic variables with earnings in the year prior to month 't', while the "Coarse III" column augments the demographic variables with the labor force status transition variables that we find to be crucial in determining participation in JTPA for all four groups. The final column presents estimates based on the "best predictor" probabilities constructed by Heckman, Ichimura, Smith and Todd (1998).[32]

The first row in Table 7 for each demographic group presents estimates from local linear matching, which assumes that Y_0 is mean independent of D conditional on $P(X)$, so there is no selection bias in any period. The second row presents estimates from a nonparametric conditional difference-in-differences estimator, which assumes that the selection bias in Y_0, conditional on $P(X)$, is the same in symmetric quarters around month 't', so that it can be differenced out.

Several important patterns emerge from the estimates in Table 7. First, for adult males, either estimation method, even if applied using the "Coarse I" probabilities based on only a few demographic variables, substantially reduces selection bias relative to the unadjusted difference in means. This indicates that conditioning on eligibility alone is easily improved on by conditioning on basic demographic variables.

Second, the strongest effect of using the probabilities of participation based

[31]The exact X for each set of participation probabilities are defined in the notes to Table 7.

[32]The X used for these best predictor probabilities are rough supersets of those used in Table 6, with some variables (*e.g.*, marital status for adult males) measured slightly differently, a small number of variables omitted (*e.g.*, family income for adult males), and a few others added. These X were iteratively selected for each demographic group based on two criteria: improvements in the prediction rate and the statistical significance of the individual variables in the participation logit. See the discussion in Appendix C of Heckman, Ichimura, Smith and Todd (1998) for more details. The slight differences between these estimates and those reported in Table 6 do not affect any of the conclusions of this paper.

on labor force histories is for adult men. For them, adding the labor force status transition pattern variables to the set of Xs used to construct the participation probabilities, which corresponds to moving from the "Coarse I" probabilities to the "Coarse III" probabilities, reduces the selection bias from -\$5,238, a figure even larger than the biases found in Table 1, to only -\$450. In contrast, for the other three demographic groups the bias from the "Coarse I" probabilities is not statistically different from that from the "Coarse III" or the "best predictor" probabilities.

Our analysis does not provide a definitive answer as to why the pattern of bias in Table 7 for the local linear matching estimator is so different for adult males compared to the other demographic groups. In addition, three factors may be at work: (1) The unadjusted mean bias for adult males is substantially larger than it is for the other three groups, for which it is not statistically distinguishable from zero. Thus, it may be that the effects of the matching for the other groups are dwarfed by the sampling variation. (2) Looking back to Table 5, the improvement in the prediction rate that results from adding the labor force status transition variables to the background variables – that is, in going from the "Coarse I" specification to the "Coarse III" specification – is around 0.11 for adult males but only about 0.07 for the other three groups. Thus, these variables may be more important in explaining participation, and controlling selection bias, for adult males. (3) The overall prediction rate for adult males in Table 5 is always about 0.10 higher than for the other three demographic groups. We conjecture that this is because family factors, which are not measured well in our data, affect participation more strongly for the other three groups than for adult males. It may be that conditioning on these other factors as well as on the labor force status transition variables is required in order to observe substantial bias reduction for groups other than adult males.

Third, for all four demographic groups, the biases resulting from all four models for the probability of participation are roughly the same for the conditional (on $P(X)$) variant of the difference-in-differences estimator, and are never statistically distinguishable from zero at conventional significance levels. For adult males, this method produces a substantial improvement over the unadjusted mean difference, and suggests that, given conditioning on at least the demographic variables in the "Coarse I" probabilities, much of the bias is constant over time in the sense of being roughly equal in symmetric intervals around the month of random assignment or measured eligibility. In contrast, for the other three demographic groups, the lack of any statistical difference between the unadjusted mean differences and the conditional on $P(X)$ difference-in-differences estimates of the bias suggests that little if any of the conditional bias is constant over time, or is equal in symmetric intervals around month 't'.

Fourth, with the exception of the adult males, for whom the conditional difference-in-differences estimator yields smaller estimated bias, there is no systematic difference in the extent of bias reduction in the two estimators considered here.

6. *Summary and Conclusions*

Using rich data on randomized-out control group members from a recent experimental evaluation of the JTPA program, we examine the earnings patterns of persons who would have participated in the counterfactual state where they do not participate. Combining the control group data with two comparison groups of persons eligible for JTPA, we consider the implications of the control (and comparison) group earnings patterns for commonly used before-after and conventional difference-in-differences estimators. In addition, we use these rich data sources to gain a deeper understanding of the determinants of participation in training programs. Six main findings emerge from our analysis.

First, a pre-program dip in the mean earnings of participants is found for all demographic groups in the JTPA data. It is a feature of the pre-program earnings of participants in many social programs (Heckman, LaLonde and Smith, 1999). Second, earnings data on the experimental control group reveal that the dip in mean earnings for participants is not mean-reverting except for adult males. In the other three groups, program participants would experience earnings growth in the post-program period even if they did not participate. This growth leads to substantial upward bias in before-after estimators of program impact.

Third, comparison groups of program eligibles exhibit different pre-program and post-program earnings patterns than experimental control group members. That is, conditioning on eligibility status for the program results in a comparison group that does not represent the desired counterfactual outcome that participants would experience if they did not participate. Using two separate comparison groups of eligibles, we show that these differences in earnings patterns lead to substantial bias in conventional difference-in-differences estimators of program impacts. Furthermore, these estimators exhibit striking instability with respect to changes in the "before" and "after" time periods used to construct them. The nonparametric conditional difference-in-differences estimators introduced in Heckman, Ichimura and Todd (1997, 1998) perform much better than the conventional difference-in-differences estimator.

Fourth, labor force status transitions, particularly transitions into unemployment from employment or from outside the labor force, drive participation in JTPA among program eligibles. Earnings changes are only weak predictors of program participation. The emphasis on earnings declines as predictors of program participation in the previous literature reflects the lack of available data in earlier studies and helps to account for the disappointing performance of some of the earlier longitudinal evaluation strategies that use lagged earnings to control for selective differences between participants and non-participants.

Our evidence suggests that the model of program participation developed by Heckman (1978) and applied by Ashenfelter and Card (1985) and others should be amended. That model emphasizes changes in the opportunity costs of earnings foregone as the major determinant of participation in training programs. The evidence suggests that it is changes in labor force (and for adult women marital and family) status that predict participation in programs. Heckman, LaLonde and Smith (1999) develop a model of labor force dynamics and training that extends the original Heckman model to account for the lessons of this paper.

Fifth, based on our finding of the importance of labor force status dynamics in determining participation in JTPA training, we investigate the performance of two estimators - the method of matching and a nonparametric conditional difference-in-differences estimator introduced in Heckman, Ichimura and Todd (1997, 1998). Especially for adult males, we find that both local linear matching and the conditional (on the probability of participation) nonparametric version of the method of difference-in-differences substantially reduce the extent of selection bias in non-experimental estimates of the impact of training on earnings. Furthermore, we find that for adult males, but not for the other three demographic groups, conditioning on labor force status transitions plays an important role in reducing the level of selection bias.

Sixth, the methods used in this paper are based on selection on observables in the sense of Heckman and Robb (1985, 1986). They reduce, but do not eliminate selection bias. The nonparametric selection bias estimator proposed and implemented in Heckman, Ichimura, Smith and Todd (1999) does not assume selection on observables. It is a promising candidate for investigation in future non-experimental evaluations of social programs.

Bibliography

[1] Amemiya, Takeshi. (1985). *Advanced Econometrics.* Cambridge, MA: Harvard University Press.

[2] Ashenfelter, Orley. (1978). 'Estimating the Effect of Training Programs on Earnings.' *Review of Economics and Statistics*, vol. 60, pp. 47-57.

[3] Ashenfelter, Orley, and Card, David. (1985). 'Using the Longitudinal Structure of Earnings to Estimate the Effect of Training Programs.' *The Review of Economics and Statistics*, vol. 67, pp. 648-660.

[4] Barnow, Burt. (1993). 'Getting It Right: Thirty Years of Changing Federal, State, and Local Relationships in Employment and Training Programs.' *Publius: The Journal of Federalism*, vol. 23, pp. 75-91.

[5] Bassi, Laurie. (1983). 'The Effect of CETA on the Post-Program Earnings of Participants.' *Journal of Human Resources*, vol. 18, pp. 539-556.

[6] Bassi, Laurie. (1984). 'Estimating the Effect of Training Programs with Non-Random Selection.' *The Review of Economics and Statistics*, vol.66, pp. 36-43.

[7] Bloom, Howard. (1991). *The National JTPA Study: Baseline Characteristics of the Experimental Sample.* Bethesda, MD: Abt Associates.

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

[9] Bloom, Howard, Orr, Larry, Bell, Stephen, Cave, George, Doolittle, Fred, Lin, Winston and Bos, Johannes. (1997). 'The Benefits and Costs of JTPA Title II-A Programs.' *Journal of Human Resources*, vol. 32, pp. 549-576.

[10] Bryant, E. and Rupp, Kalman. (1987). 'Evaluating the Impact of CETA on Participant Earnings.' *Evaluation Review*, vol. 11, pp. 473-492.

[11] Campbell, Donald and Stanley, Julian. (1966). Experimental and Quasi-Experimental Designs for Research. Chicago, IL: Rand McNally.

[12] Devine, Theresa and Heckman, James. (1996). 'The Structure and Consequences of Eligibility Rules for a Social Program.' In *Research in Labor Economics, Volume 15* (Solomon Polachek, ed.), pp. 111-170. Greenwich, CT: JAI Press.

[13] Dickinson, K., Johnson, Terry and West, Rebecca. (1987). 'An Analysis of the Sensitivity of Quasi-experimental Net Impact Estimates of CETA Programs.' *Evaluation Review*, vol. 11, pp. 452-472.

[14] Heckman, James. (1978). 'Dummy Endogenous Variables in a Simultaneous Equations System.' *Econometrica*, vol. 46, pp. 931-59.

[15] Heckman, James. (1979). 'Sample Selection Bias as a Specification Error.' *Econometrica*, vol. 47, pp. 153-161.

[16] Heckman, James. (1992). 'Randomization and Social Program Evaluation.' In *Evaluating Welfare and Training Programs* (Charles Manski and Irwin Garfinkel,eds.), pp. 201-230. Cambridge, MA: Harvard University Press.

[17] Heckman, James. (1997). 'Instrumental Variables: A Study of Implicit Behavioral Assumptions Used in Making Program Evaluations.' *Journal of Human Resources*, vol. 32, pp. 441-462.

[18] Heckman, James, Hohmann, Neil, and Smith, Jeffrey, with Khoo, Michael. (1998). 'Substitution and Dropout Bias in Social Experiments: A Study of an Influential Social Experiment.' University of Chicago, Unpublished manuscript.

[19] Heckman, James, Ichimura, Hidehiko, Smith, Jeffrey and Todd, Petra. (1996). 'Sources of Selection Bias in Evaluating Social Programs: An Interpretation of Conventional Measures and Evidence on the Effectiveness of Matching as a Program Evaluation Method.' *Proceedings of the National Academy of Sciences*, vol. 93, pp. 13416-13420.

[20] Heckman, James, Ichimura, Hidehiko, Smith, Jeffrey and Todd, Petra. (1998). 'Characterizing Selection Bias Using Experimental Data.' *Econometrica*, vol. 66, pp. 1017-1098.

[21] Heckman, James, Ichimura, Hidehiko and Todd, Petra. (1997). 'Matching as an Econometric Evaluation Estimator: Evidence from Evaluating a Job Training Program.' *Review of Economic Studies*, vol. 64, pp. 605-654.

[22] Heckman, James, Ichimura, Hidehiko and Todd, Petra. (1998). 'Matching as an Econometric Evaluation Estimator.' *Review of Economic Studies*, vol. 65, pp. 261-294.

[23] Heckman, James, Khoo, Michael, Roselius, Rebecca and Smith, Jeffrey. (1996). 'The Empirical Importance of Randomization Bias in Social Experiments: Evidence from the National JTPA Study.' University of Chicago, unpublished manuscript.

[24] Heckman, James and Smith, Jeffrey. (1993). 'Assessing the Case for Randomized Evaluation of Social Programs.' In *Measuring Labour Market Measures: Evaluating the Effects of Active Labour Market Policies* (Karsten Jensen and Per Kongshoj Madsen, eds.), pp. 35-96. Copenhagen: Danish Ministry of Labour.

[25] Heckman, James and Smith, Jeffrey. (1995). 'Assessing the Case for Social Experiments.' *Journal of Economic Perspectives*, vol. 9, pp. 85-110.

[26] Heckman, James, LaLonde, Robert and Smith, Jeffrey. (1999). 'The Economics and Econometrics of Training Programs.' In *Handbook of Labor Economics, Volumes 3 and 4* (Orley Ashenfelter and David Card, eds.), forthcoming. Amsterdam: North-Holland.

[27] Heckman, James and Robb, Richard. (1985). 'Alternative Methods for Evaluating the Impact of Interventions.' In *Longitudinal Analysis of Labor Market Data* (James Heckman and Burton Singer, eds.), pp. 156-245. Cambridge: Cambridge University Press.

[28] Heckman, James and Robb, Richard. (1986). 'Alternative Methods For Solving The Problem of Selection Bias In Evaluating The Impact of Treatments

on Outcomes.' In Drawing Inference From Self-Selected Samples (Howard Wainer, ed.), pp. 63-107. Berlin: Springer-Verlag.

[29] Heckman, James and Roselius, Rebecca. (1994a). 'Evaluating the Impact of Training on the Earnings and Labor Force Status of Young Women: Better Data Help A Lot.' University of Chicago, Unpublished manuscript.

[30] Heckman, James and Roselius, Rebecca. (1994b). 'Nonexperimental Evaluation of Job Training Programs for Young Men.' University of Chicago, Unpublished manuscript.

[31] Heckman, James and Smith, Jeffrey. (1997). 'The Determinants of Participation in a Social Program: Evidence from JTPA.' University of Chicago, Unpublished manuscript.

[32] Heckman, James and Smith, Jeffrey. (1998). 'Evaluating the Welfare State.' In *Econometrics and Economic Theory in the 20th Century: The Ragnar Frisch Centennial* (Steiner Strom, ed.), forthcoming. Cambridge: Cambridge University Press for the Econometric Society Monograph Series.

[33] Heckman, James and Todd, Petra. (1996). 'Assessing the Performance of Alternative Estimators of Program Impacts: A Study of Adult Men and Women in JTPA.' University of Chicago, Unpublished manuscript.

[34] Hunt, Allen; Rupp, Kalman and Associates. (1984). 'The Implementation of Title II-A of JTPA in the States and Service Delivery Areas: The New Partnership and Program Directions.' *Proceedings of the Thirty-Seventh Annual Meeting of the Industrial Relations Research Association*, pp. 85-92.

[35] LaLonde, Robert. (1986). 'Evaluating the Econometric Evaluations of Training Programs with Experimental Data.' *American Economic Review*, vol. 76, pp. 604-620.

[36] Rosenbaum, Paul and Rubin, Donald. (1983). 'The Central Role of the Propensity Score in Observational Studies for Causal Effects.' *Biometrika*, vol. 70, pp. 41-55.

[37] Sandell, Steven and Rupp, Kalman. (1988). 'Who is Served in JTPA Programs: Patterns of Participation and Intergroup Equity.' Research Report RR-88-03, National Commission for Employment Policy.

[38] Smith, Jeffrey. (1997a). 'Measuring Earnings Dynamics Among the Poor: A Comparison of Two Samples of JTPA Eligibles.' University of Western Ontario, Unpublished manuscript.

[39] Smith, Jeffrey. (1997b). 'Measuring Earnings Levels Among the Poor: A Comparison of Two Samples of JTPA Eligibles.' University of Western Ontario, Unpublished manuscript.

[40] U.S. Department of Labor. (1993a). *Job Training Quarterly Survey: JTPA Title IIA and III Enrollments and Terminations During Program Year 1991 (July 1991-June 1992).* Employment and Training Administration, Office of Strategic Planning and Policy Development, Division of Performance Management and Evaluation.

[41] U.S. Department of Labor. (1993b). *Title II Eligibility Documentation.* Employment and Training Administration.

TABLE 1

BEFORE-AFTER AND DIFFERENCE-IN-DIFFERENCES IMPACT ESTIMATES

NJS Control and Treatment Group and ENP Samples and SIPP Eligible Sample
Estimates in Dollars of Earnings Per 18 Months
Estimated Standard Errors in Parentheses

Before Period	After Period	Experimental Estimate [1]	Before-After Estimate [2]	Diff-in-Diffs Estimate (ENP) [3]	Diff-in-Diffs Estimate (SIPP) [3]
Adult Males					
t-18 to t-1	t+1 to t+18	656.93 (562.78)	3108.98 (511.25)	922.89 (1143.08)	N.A.
t-15 to t-1	t+1 to t+15	601.51 (578.83)	3413.89 (512.78)	1501.74 (1144.06)	851.26 (559.09)
t-12 to t-1	t+1 to t+12	529.75 (586.53)	3592.49 (512.52)	1994.42 (1141.67)	1163.95 (544.36)
t-15 to t-13	t+13 to t+15	953.74 (703.98)	2626.14 (693.13)	-807.58 (1351.88)	-347.21 (748.59)
t-18 to t-16	t+18 to t+16	937.18 (687.96)	1854.91 (697.06)	-1960.56 (1342.43)	N.A.
Adult Females					
t-18 to t-1	t+1 to t+18	845.17 (370.54)	3171.83 (317.11)	2188.61 (536.84)	N.A.
t-15 to t-1	t+1 to t+15	877.88 (368.53)	3078.08 (315.99)	2236.26 (537.80)	2059.30 (329.08)
t-12 to t-1	t+1 to t+12	928.02 (366.41)	2986.58 (313.45)	2269.11 (537.48)	2106.32 (322.75)
t-15 to t-13	t+13 to t+15	753.15 (481.93)	3413.24 (425.47)	1989.68 (632.21)	1870.15 (440.78)
t-18 to t-16	t+18 to t+16	811.95 (476.98)	3600.43 (424.58)	1878.65 (631.71)	N.A.

1. Experimental estimates include only treatments and controls at the four training centers at which detailed information on controls and eligibles non-participants was collected. The experimental estimates presented are cross-section estimates obtained by differencing treatment and control mean earnings in the "after" period. Note that with experimental data the expected values of the cross section and difference-in-differences estimators are the same.
2. The before-after estimates are obtained by subtracting control group mean earnings in the "before" period from treatment group mean earnings in the "after" period.
3. The difference-in-differences estimates consist of the difference between the change in mean earnings for the treatment group between the "before" and "after" periods and the change in mean earnings for the comparison group (either SIPP or ENP) between the "before" and "after" periods. Treatment group mean earnings in the "before" period are estimated using control group mean earnings in the "before" period.
4. Some values for the SIPP are omitted due to the limited length of the panel.
5. The top one percent of monthly earnings are trimmed for each demographic group in each of the SIPP, ENP, control and treatment group samples.
6. The control group and ENP samples include only persons with a valid earnings observation in month t-18 and in month t+18. The treatment group sample includes only persons with a valid earnings observation for month t+18 (no earnings information is available for the treatment group prior to month t). The SIPP eligible sample includes only persons with valid earnings information in the first and last months of the panel.
7. Sample sizes are 1271 treatments, 453 controls, 401 ENPs, and 10864 SIPP eligibles for adult males and 1464 treatments, 599 controls, 885 ENPs and 19606 SIPP eligibles for adult females. The SIPP eligible sample consists of person-months rather than persons.

TABLE 1 (CONTINUED)

BEFORE-AFTER AND DIFFERENCE-IN-DIFFERENCES IMPACT ESTIMATES

NJS Control and Treatment Group and ENP Samples and SIPP Eligible Sample
Estimates in Dollars of Earnings Per 18 Months
Estimated Standard Errors in Parentheses

Before Period	After Period	Experimental Estimate [1]	Before-After Estimate [2]	Diff-in-Diffs Estimate (ENP) [3]	Diff-in-Diffs Estimate (SIPP) [3]
Male Youth					
t-18 to t-1	t+1 to t+18	-1060.02 (658.69)	2498.51 (527.17)	-1214.98 (1404.83)	N.A.
t-15 to t-1	t+1 to t+15	-1134.28 (690.22)	2014.13 (540.16)	-1156.27 (1430.55)	349.10 (581.61)
t-12 to t-1	t+1 to t+12	-1035.84 (694.07)	1486.05 (559.63)	-978.22 (1472.85)	-154.12 (592.99)
t-15 to t-13	t+13 to t+15	-1500.32 (884.89)	4665.60 (647.23)	-2065.06 (1675.42)	1855.96 (701.56)
t-18 to t-16	t+18 to t+16	-1387.61 (843.04)	4836.80 (667.81)	-1957.95 (1756.79)	N.A.
Female Youth					
t-18 to t-1	t+1 to t+18	-112.83 (432.68)	2641.68 (322.33)	1180.19 (742.50)	N.A.
t-15 to t-1	t+1 to t+15	-96.70 (440.35)	2452.41 (328.15)	1313.47 (754.56)	1031.85 (359.06)
t-12 to t-1	t+1 to t+12	-148.89 (445.28)	2223.29 (333.17)	1327.33 (773.33)	904.51 (355.72)
t-15 to t-13	t+13 to t+15	292.14 (583.40)	3492.87 (435.97)	1302.14 (885.76)	1163.15 (478.07)
t-18 to t-16	t+18 to t+16	-105.99 (603.27)	3683.81 (427.29)	490.73 (918.57)	N.A.

1. Experimental estimates include only treatments and controls at the four training centers at which detailed information on controls and eligibles non-participants was collected. The experimental estimates presented are cross-section estimates obtained by differencing treatment and control mean earnings in the "after" period. Note that with experimental data the expected values of the cross section and difference-in-differences estimators are the same.
2. The before-after estimates are obtained by subtracting control group mean earnings in the "before" period from treatment group mean earnings in the "after" period.
3. The difference-in-differences estimates consist of the difference between the change in mean earnings for the treatment group between the "before" and "after" periods and the change in mean earnings for the comparison group (either SIPP or ENP) between the "before" and "after" periods. Treatment group mean earnings in the "before" period are estimated using control group mean earnings in the "before" period.
4. Some values for the SIPP are omitted due to the limited length of the panel.
5. The top one percent of monthly earnings are trimmed for each demographic group in each of the SIPP, ENP, control and treatment group samples.
6. The control group and ENP samples include only persons with a valid earnings observation in month t-18 and in month t+18. The treatment group sample includes only persons with a valid earnings observation for month t+18 (no earnings information is available for the treatment group prior to month t). The SIPP eligible sample includes only persons with valid earnings information in the first and last months of the panel.
7. Sample sizes are 736 treatments, 230 controls, 85 ENPs and 2167 SIPP eligibles for male youth and 804 treatments, 289 controls, 154 ENPs and 3311 SIPP eligibles for female youth. The SIPP eligible sample consists of person-months rather than persons.

TABLE 2
RATES OF PARTICIPATION IN JTPA CONDITIONAL ON ELIGIBILITY BY LABOR FORCE STATUS AND LABOR FORCE STATUS TRANSITION
NJS ENP and Control Group Samples

	Adult Males	Adult Females	Male Youth	Female Youth
Labor Force Status at ‘t’				
Employed	0.0137	0.0197	0.0221	0.0204
	(0.0007)	(0.0010)	(0.0019)	(0.0021)
Unemployed	0.1171	0.1017	0.0484	0.0868
	(0.0106)	(0.0056)	(0.0068)	(0.0109)
OLF	0.0392	0.0197	0.0300	0.0201
	(0.0041)	(0.0010)	(0.0053)	(0.0017)
Labor Force Status Transitions				
Emp -> Emp	0.0084	0.0140	0.0166	0.0115
	(0.0007)	(0.0011)	(0.0019)	(0.0017)
Unm -> Emp	0.0496	0.0483	0.0615	0.0444
	(0.0080)	(0.0074)	(0.0163)	(0.0131)
OLF -> Emp	0.0551	0.0269	0.0228	0.0316
	(0.0122)	(0.0048)	(0.0053)	(0.0082)
Emp -> Unm	0.1433	0.1330	0.0631	0.1446
	(0.0165)	(0.0120)	(0.0128)	(0.0447)
Unm -> Unm	0.0967	0.0948	0.0333	0.0631
	(0.0142)	(0.0100)	(0.0097)	(0.0159)
OLF -> Unm	0.1182	0.0693	0.0400	0.0725
	(0.0377)	(0.0089)	(0.0161)	(0.0206)
Emp -> OLF	0.1032	0.0355	0.0713	0.0332
	(0.0268)	(0.0049)	(0.0385)	(0.0071)
Unm -> OLF	0.1363	0.0500	0.0289	0.0240
	(0.0171)	(0.0107)	(0.0181)	(0.0106)
OLF -> OLF	0.0275	0.0166	0.0146	0.0155
	(0.0040)	(0.0011)	(0.0051)	(0.0020)

1. The entries in the table are the conditional percent participating calculated using the controls and ENPs under the assumption that the population participation rate is 0.03.
2. Labor force status transitions are defined by looking backward in time starting in month ‘t’ and ending in month ‘t-6’. The second status in each pattern is the status in month ‘t’. The first status is the most recent prior status within the indicated time period. Thus, "Emp -> Unm" indicates persons unemployed at ‘t’ but whose most recent preceding labor force status within the prior six months was employed. Repeated patterns such as "Emp -> Emp" indicate persons with the same labor force status from ‘t-6’ to ‘t’.

TABLE 3

DATA AVAILABLE IN STUDIES OF EMPLOYMENT AND TRAINING PROGRAMS

	Ashenfelter (1978)	Ashenfelter and Card (1985)	LaLonde (1986)	Bryant and Rupp (1987)	Dickinson, Johnson and West (1987)	NJS
Demographic Variables						
Age	Yes	Yes	Yes	Yes	Yes	Yes
Sex	Yes	Yes	Yes	Yes	Yes	Yes
Race or ethnicity	Yes	Yes	Yes	Yes	Yes	Yes
Years of schooling	No	Yes	Yes	Yes	Yes	Yes
Marital status	No	Yes	Yes	Yes	Yes	Yes
Transfer Program Participation Variables						
AFDC receipt	No	No	No	No	No	Yes
Food stamp receipt	No	No	No	No	No	Yes
Labor Market Variables						
Pre-training hours	No	No	2 Years (Annual)	No	No	5 Years (Monthly)
Post-training hours	No	No	2 Years (Annual)	No	No	2 Years (Monthly)
Pre-training earnings and employment	5 Years (Annual)	5 Years (Annual)	2 Years (Annual)	2 Years (Annual)	2 Years (Annual)	5 Years (Monthly)
Post-training earnings and employment	5 Years (Annual)	2 Years (Annual)	2 Years (Annual)	2 Years (Annual)	2 Years (Annual)	2 Years (Monthly)
Pre-training labor force status	No	No	No	No	No	1 Year (Monthly)
Post-training labor force status	No	No	No	No	No	2 Years (Monthly)
Local labor market	No	No	No	No	No	Yes
Other Variables						
Family income	No	No	No	No	No	Yes

1. NJS refers to studies based on the National JTPA Study data. In addition to this paper, these include Heckman, Ichimura, Smith and Todd (1998), Heckman and Smith (1997), Heckman and Todd (1996) and Heckman and Roselius (1994a,b), among others.

2. Ashenfelter (1978), Ashenfelter and Card (1985), Bryant and Rupp (1987) and Dickinson, Johnson and West (1987) had Social Security earnings data matched to samples of program participants and to comparison groups constructed from the Current Population Survey. Lalonde had self-reported data on Supported Work experimental treatment and control group members, along with self-reported data on PSID sample members and Social Security earnings data on CPS comparison group members. The NJS studies have available constructed monthly earnings measures based on self-reported information about job spells for experimental control group members and for a comparison group of eligible non-participants at four of the 16 training centers in the study.

TABLE 4
DEFINITIONS OF LABOR MARKET MEASURES

Background Specifications

1. The background (BKGD) specification includes race and ethnicity indicators, age category indicators, years of completed schooling category indicators, marital status indicators, and an indicator for the presence of a child less than six years of age. These variables are included in most of the other specifications as well.

2. The family income specification adds a categorical measure of family income based on the earnings in the 12 months prior to the baseline interview of all family members living in the same household as the sample member at the time of the interview.

Employment Measures

1. "Employment at 't'" is an indicator for whether or not the person was employed in the month of random assignment or eligibility determination.

2. "Employment transitions" are the four patterns formed by the employment status at 't' and the most recent previous employment status in months 't-1' to 't-6'. The patterns are continuously employed, job loser, job gainer and continuously not employed.

3. "ETC" indicates categories of the number of transitions from employment to non-employment or vice versa in the 24 months prior to the baseline interview.

4. "18 month job spells" indicates categories of the total number of job spells in the 18 months prior to random assignment or eligibility determination.

5. "48 month job spells" indicates categories of the total number of job spells in the 48 months prior to random assignment or eligibility determination.

Earnings Measures

1. "Earnings in 't-1' to 't-6'" are own total earnings in each of the six months prior to random assignment or eligibility determination.

2. "Earnings in 'Q-1' to 'Q-4'" are own total earnings in each of the four quarters prior to random assignment or eligibility determination.

Labor Force Status Measures

1. "LFS at 't'" is the labor force status (employed, unemployed, or out of the labor force) in the month of random assignment or eligibility determination.

2. Time in labor force status is the number of months in the labor force status at random assignment or eligibility determination. There are separate variables for each status: employed, unemployed, and out of the labor force. For each status, there is a continuous variable for 0-6 months and an indicator variable for greater than six months in the status.

3. "2 Quarter LFS" consists of patterns formed by constructing quarterly labor force measures for the two quarters prior to random assignment or eligibility determination. That is, the statuses are first aggregated within quarters, with employment having precedence over unemployment and unemployment over OLF, and then combined into one of nine possible sequences.

4. "6 Month LFS2" is the two most recent labor force statuses in the seven months up to and including the month of random assignment or eligibility determination as defined in the text.

TABLE 5

JTPA PARTICIPATION PROBABILITY EQUATIONS

PREDICTION CUTOFF VALUE = 0.03

Mean Percent Correctly Predicted
Estimated Standard Errors in Parentheses
NJS ENP and Control Samples

Specification	Adult Males	Adult Females	Male Youth	Female Youth
Background				
Background (BKGD)	.7010	.6362	.5968	.6172
	(.0116)	(.0100)	(.0226)	(.0185)
BKGD + Family income	.7453	.6317	.6372	.6240
	(.0111)	(.0101)	(.0219)	(.0185)
Employment Specifications				
Employment at 't' (No BKGD)	.7457	.5842	.5868	.5638
	(.0110)	(.0099)	(.0226)	(.0179)
BKGD + Employment at 't'	.7664	.6543	.6295	.6361
	(.0108)	(.0100)	(.0223)	(.0184)
BKGD + Employment transition	.8043	.6650	.6387	.6756
	(.0101)	(.0100)	(.0220)	(.0179)
BKGD + ETC	.7700	.6761	.6779	.6599
	(.0107)	(.0098)	(.0216)	(.0181)
BKGD + 18 month job spells	.7129	.6390	.6384	.6331
	(.0115)	(.0100)	(.0217)	(.0184)
BKGD + 48 month job spells	.7086	.6632	.6110	.6427
	(.0116)	(.0100)	(.0224)	(.0183)
Earnings Specifications				
BKGD + Earnings in 't-1' to 't-6'	.7901	.6589	.6308	.6124
	(.0103)	(.0099)	(.0222)	(.0186)
BKGD + Earnings in 'Q-1' to 'Q-4'	.7933	.6464	.6152	.5986
	(.0102)	(.0100)	(.0224)	(.0187)
Labor Force Status Specifications				
LFS at 't' (No BKGD)	.7542	.6831	.6008	.6632
	(.0109)	(.0093)	(.0225)	(.0159)
BKGD + LFS at 't'	.7714	.6960	.6393	.6365
	(.0107)	(.0098)	(.0221)	(.0178)
BKGD + Time in labor force status	.8016	.7049	.6417	.6954
	(.0101)	(.0097)	(.0220)	(.0174)
BKGD + 2 quarter LFS	.7517	.6611	.6241	.6536
	(.0110)	(.0100)	(.0223)	(.0182)
BKGD + 6 month LFS2	.8104	.7003	.6724	.6878
	(.0100)	(.0098)	(.0213)	(.0176)

1. BKGD includes race, age, years of schooling, marital status, and presence of a child less than six years of age.

TABLE 6

JTPA PARTICIPATION PROBABILITY ESTIMATES

Weighted Logit Equation - Dependent Variable: Control Status
NJS ENP and Control Samples

	Adult Males			Adult Females		
Variable	**Coefficient**	**Standard Error**	**Numerical Derivative**	**Coefficient**	**Standard Error**	**Numerical Derivative**
Black	0.149	0.273	0.004	0.234	0.174	0.006
Hispanic	-0.256	0.308	-0.006	0.345	0.192	0.009
Other race-ethnic	0.394	0.409	0.011	0.262	0.337	0.007
Age 30-39	-0.458	0.228	-0.012	-0.294	0.139	-0.008
Age 40-49	-0.982	0.297	-0.022	-0.229	0.180	-0.006
Age 50-54	-0.400	0.394	-0.011	-0.349	0.281	-0.009
Highest grade < 10	-0.541	0.272	-0.012	-0.498	0.154	-0.012
Highest grade 10-11	0.354	0.266	0.010	-0.063	0.162	-0.002
Highest grade 13-15	0.711	0.311	0.023	0.168	0.201	0.005
Highest grade > 15	-1.373	0.416	-0.022	-0.414	0.389	-0.011
Currently married	-0.522	0.263	-0.012	-0.904	0.184	-0.019
Married 1-24 months ago	-0.029	0.637	-0.001	0.564	0.225	0.021
Married > 24 months ago	1.240	0.487	0.052	1.263	0.199	0.064
Child age < 6 years	-0.217	0.311	-0.005	-0.245	0.133	-0.006
Received AFDC at ‘t’	-1.196	0.511	-0.020	-0.758	0.205	-0.019
Received food stamps at ‘t’	0.560	0.244	0.015	0.452	0.174	0.013
Unemployed -> Employed	1.927	0.322	0.043	1.556	0.276	0.034
OLF -> Employed	2.083	0.418	0.051	0.988	0.320	0.016
Employed -> Unemployed	3.239	0.330	0.136	2.825	0.244	0.121
Unemployed -> Unemployed	2.766	0.412	0.094	2.621	0.263	0.101
OLF -> Unemployed	3.787	0.490	0.199	2.146	0.297	0.065
Employed -> OLF	2.684	0.626	0.087	1.214	0.271	0.022
Unemployed -> OLF	3.633	0.774	0.179	1.991	0.339	0.055
OLF -> OLF	1.186	0.411	0.016	0.770	0.222	0.011
Family Income $3K-9K	-0.347	0.364	-0.011	0.543	0.207	0.016
Family Income $9K-15K	-0.025	0.400	-0.001	0.162	0.284	0.004
Family Income > $15K	-1.599	0.474	-0.034	0.029	0.269	0.001

1. Number of observations: 1552 adult males and 2438 adult females.

2. The logit model also includes training center indicators and a constant.

3. The omitted race group is whites, the omitted age group is age 22-29, the omitted highest grade completed category is exactly 12 years, the omitted marital status category is never married, the omitted labor force status transition pattern is "Employed -> Employed" and the omitted family income category is less than $3000.

4. Labor force status transition patterns are defined by looking backward in time starting in month ‘t’ and ending in month ‘t-6’. The second status in each pattern is the status in month ‘t’. The first status is the most recent prior status within the indicated time period. Thus "Employed -> Unemployed" indicates a person unemployed at ‘t’ but whose most recent labor force status within the prior six months was employed.

TABLE 6 (CONTINUED)

JTPA PARTICIPATION PROBABILITY ESTIMATES

Weighted Logit Equation - Dependent Variable: Control Status
NJS ENP and Control Samples

	Male Youth			Female Youth		
Variable	**Coefficient**	**Standard Error**	**Numerical Derivative**	**Coefficient**	**Standard Error**	**Numerical Derivative**
Black	0.410	0.384	0.014	0.742	0.303	0.021
Hispanic	-0.494	0.492	-0.011	0.457	0.351	0.011
Other race-ethnic	-1.846	0.899	-0.028	-1.058	0.650	-0.014
Age 19-21	0.153	0.347	0.004	-0.535	0.265	-0.016
Highest grade < 10	0.589	0.441	0.015	-0.394	0.324	-0.011
Highest grade 10-11	0.673	0.385	0.018	-0.235	0.351	-0.007
Highest grade > 12	-0.164	0.598	-0.003	0.084	0.365	0.003
Currently married	0.298	0.461	0.009	-0.563	0.346	-0.013
Div-Wid-Sep	-0.439	0.817	-0.010	0.241	0.403	0.008
Child age < 6 years	-1.059	0.498	-0.021	-0.241	0.260	-0.007
Received AFDC at 't'	-0.721	0.730	-0.015	-0.988	0.363	-0.025
Received food stamps at 't'	-0.046	0.441	-0.001	1.363	0.337	0.052
Unemployed -> Employed	2.125	0.482	0.089	1.599	0.466	0.030
OLF -> Employed	-0.166	0.511	-0.002	1.394	0.449	0.023
Employed -> Unemployed	1.593	0.442	0.051	3.218	0.472	0.149
Unemployed -> Unemployed	1.162	0.579	0.030	2.379	0.473	0.070
OLF -> Unemployed	1.000	0.597	0.024	3.018	0.509	0.126
Employed -> OLF	2.016	0.633	0.080	1.554	0.421	0.028
Unemployed -> OLF	0.993	0.774	0.023	1.106	0.725	0.016
OLF -> OLF	0.191	0.567	0.003	0.709	0.414	0.008
Family Income \$3K-9K	1.450	0.574	0.058	-0.466	0.452	-0.010
Family Income \$9K-15K	0.110	0.602	0.002	0.031	0.528	0.001
Family Income > \$15K	0.048	0.642	0.001	1.423	0.441	0.068

1. Number of observations: 530 male youth and 701 female youth.
2. The logit model also includes training center indicators and a constant.
3. The omitted race group is whites, the omitted age group is age 16-18, the omitted highest grade completed category is exactly 12 years, the omitted marital status category is never married, the omitted labor force status transition pattern is "Employed -> Employed" and the omitted family income category is less than $3000.
4. Labor force status transition patterns are defined by looking backward in time starting in month 't' and ending in month 't-6'. The second status in each pattern is the status in month 't'. The first status is the most recent prior status within the indicated time period. Thus "Employed -> Unemployed" indicates a person unemployed at 't' but whose most recent labor force status within the prior six months was employed.

TABLE 7

ESTIMATED SELECTION BIAS IN NONEXPERIMENTAL ESTIMATES OF THE IMPACT OF JTPA TRAINING ON EARNINGS IN THE 18 MONTHS AFTER RANDOM ASSIGNMENT USING ALTERNATIVE ESTIMATORS

Estimated Standard Errors in Parentheses
NJS ENP and Control Samples

	Unadjusted Mean Difference	Coarse I Probabilities	Coarse II Probabilities	Coarse III Probabilities	"Best Predictor" Probabilities
Adult Males					
Local linear matching	-6066	-5238	-2988	-450	684
	(846)	(972)	(1008)	(1494)	(1152)
Conditional difference-in-differences	-6066	576	2592	774	936
	(846)	(1404)	(1098)	(1710)	(1332)
Adult Females					
Local linear matching	594	198	396	576	720
	(468)	(558)	(522)	(630)	(684)
Conditional difference-in-differences	594	306	36	414	486
	(468)	(558)	(540)	(702)	(702)
Male Youth					
Local linear matching	360	36	144	306	126
	(1026)	(936)	(936)	(1260)	(954)
Conditional difference-in-differences	360	666	486	612	396
	(1026)	(1008)	(972)	(1062)	(864)
Female Youth					
Local linear matching	864	648	738	1116	144
	(648)	(648)	(630)	(756)	(756)
Conditional difference-in-differences	864	936	828	504	306
	(648)	(630)	(630)	(702)	(702)

1. Source: Heckman, Ichimura and Todd (1997), Tables 6(a) and 6(b).

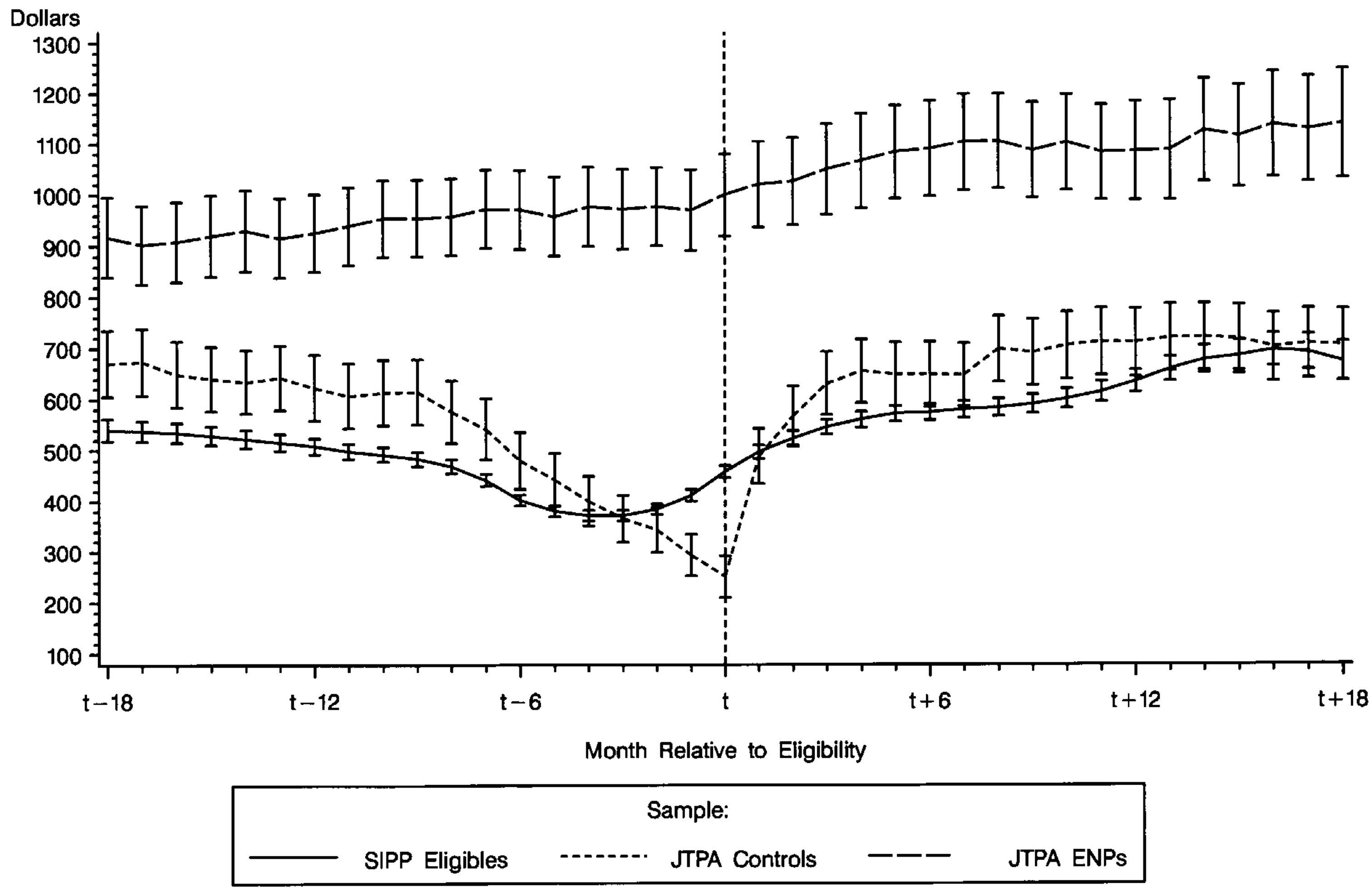
2. The best predictor probabilities for all four demographic groups contain indicator variables for training center, age, race and ethnicity, years of schooling, marital status, children less than six and labor force status transitions. The model for adult males also includes an indicator for past vocational training, the number of household members, earnings in the month of random assignment or measured eligibility (RA or EL) and indicators for the number of jobs held in the 18 months prior to RA or EL. The model for adult females includes an indicator for recent schooling, earnings in the month of RA or EL, and indicators for the number of employment transitions in the 24 months prior to RA or EL. The model for male youth includes average earnings in the six months prior to RA or EL and in the 12 months prior to RA or EL and average positive earnings in the 6 months prior to RA or EL. The model for female youth includes average earnings in the 12 months prior to RA or EL.

FIGURE 1A

MEAN SELF-REPORTED MONTHLY EARNINGS

SIPP Eligibles and JTPA Controls and ENPs

Male Adults



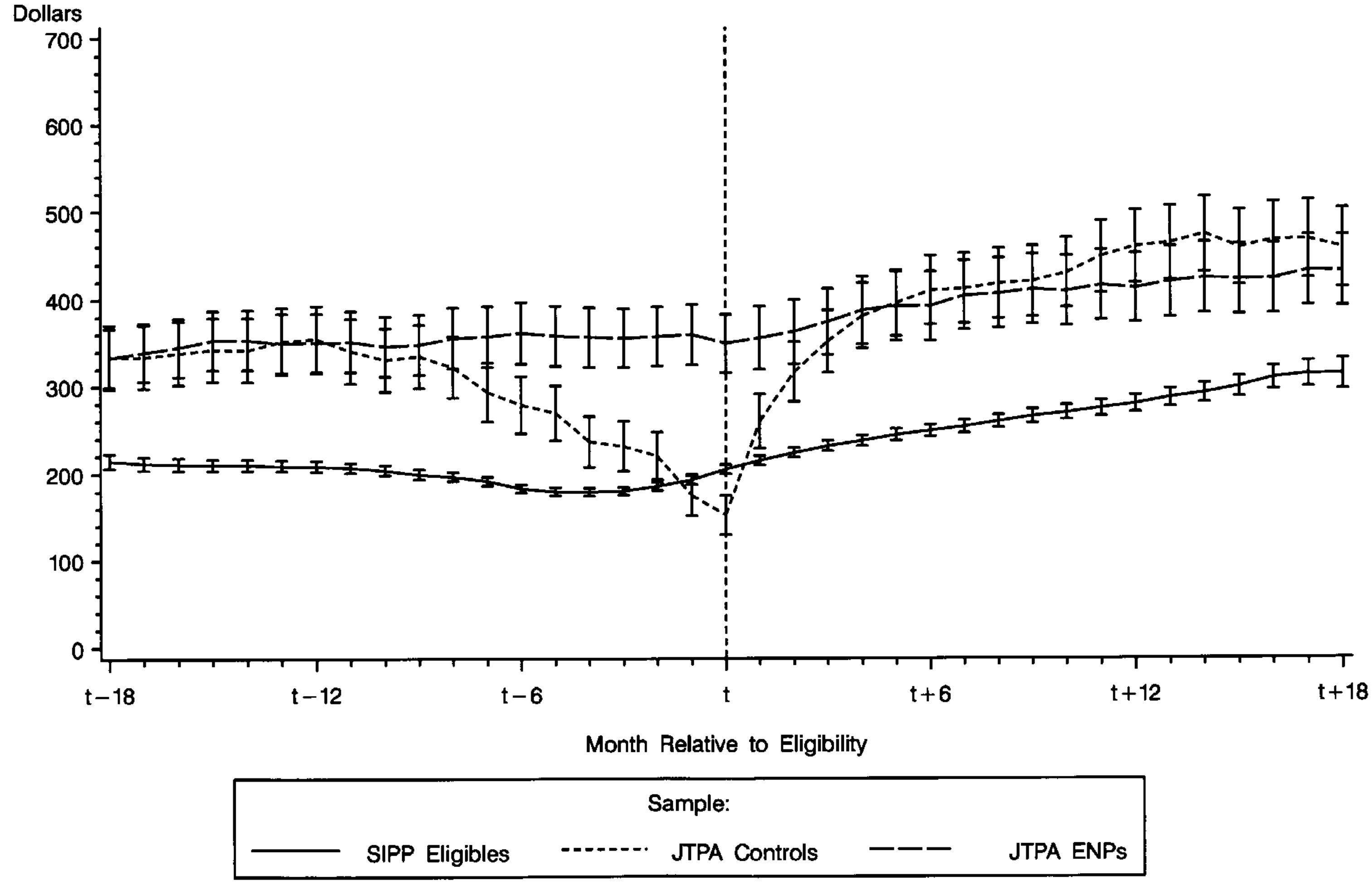
1. SIPP uses all JTPA-eligible person-month observations of respondents present in both the first and last months of the panel.
2. Controls are randomized-out participants from the National JTPA Study. Observations based on quasi-rectangular sample.
3. ENPs are JTPA-eligible non-participants at the same sites as the controls from the National JTPA Study. Observations based on quasi-rectangular sample.
4. Standard error bars +/- 2 standard errors of the means.
5. Top 1% of earnings trimmed in each month for ENP, control and SIPP samples.

FIGURE 1B

MEAN SELF-REPORTED MONTHLY EARNINGS

SIPP Eligibles and JTPA Controls and ENPs

Female Adults



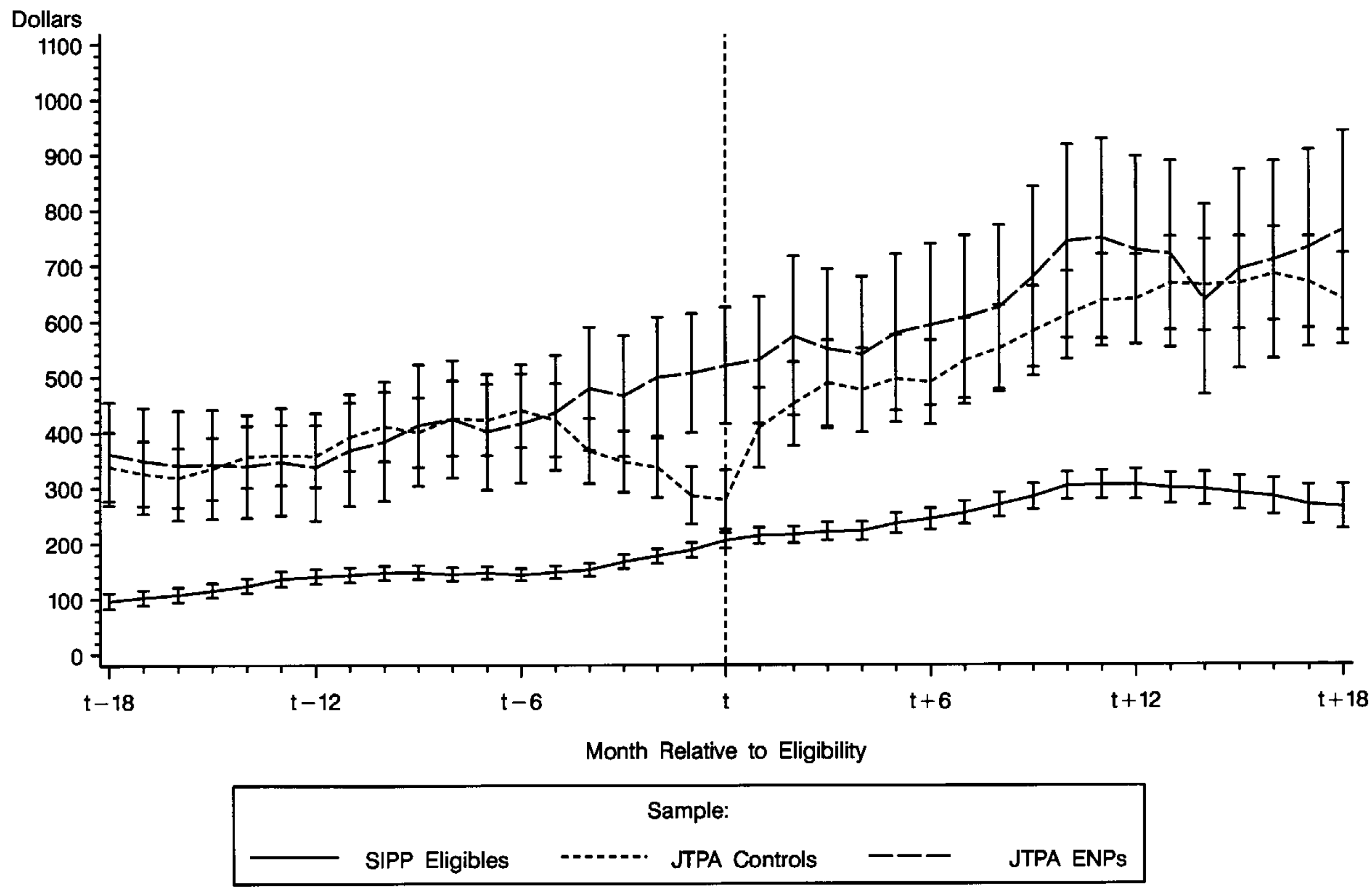
1. SIPP uses all JTPA-eligible person-month observations of respondents present in both the first and last months of the panel.
2. Controls are randomized-out participants from the National JTPA Study. Observations based on quasi-rectangular sample.
3. ENPs are JTPA-eligible non-participants at the same sites as the controls from the National JTPA Study. Observations based on quasi-rectangular sample.
4. Standard error bars +/- 2 standard errors of the means.
5. Top 1% of earnings trimmed in each month for ENP, control and SIPP samples.

FIGURE 1C

MEAN SELF-REPORTED MONTHLY EARNINGS

SIPP Eligibles and JTPA Controls and ENPs

Male Youths



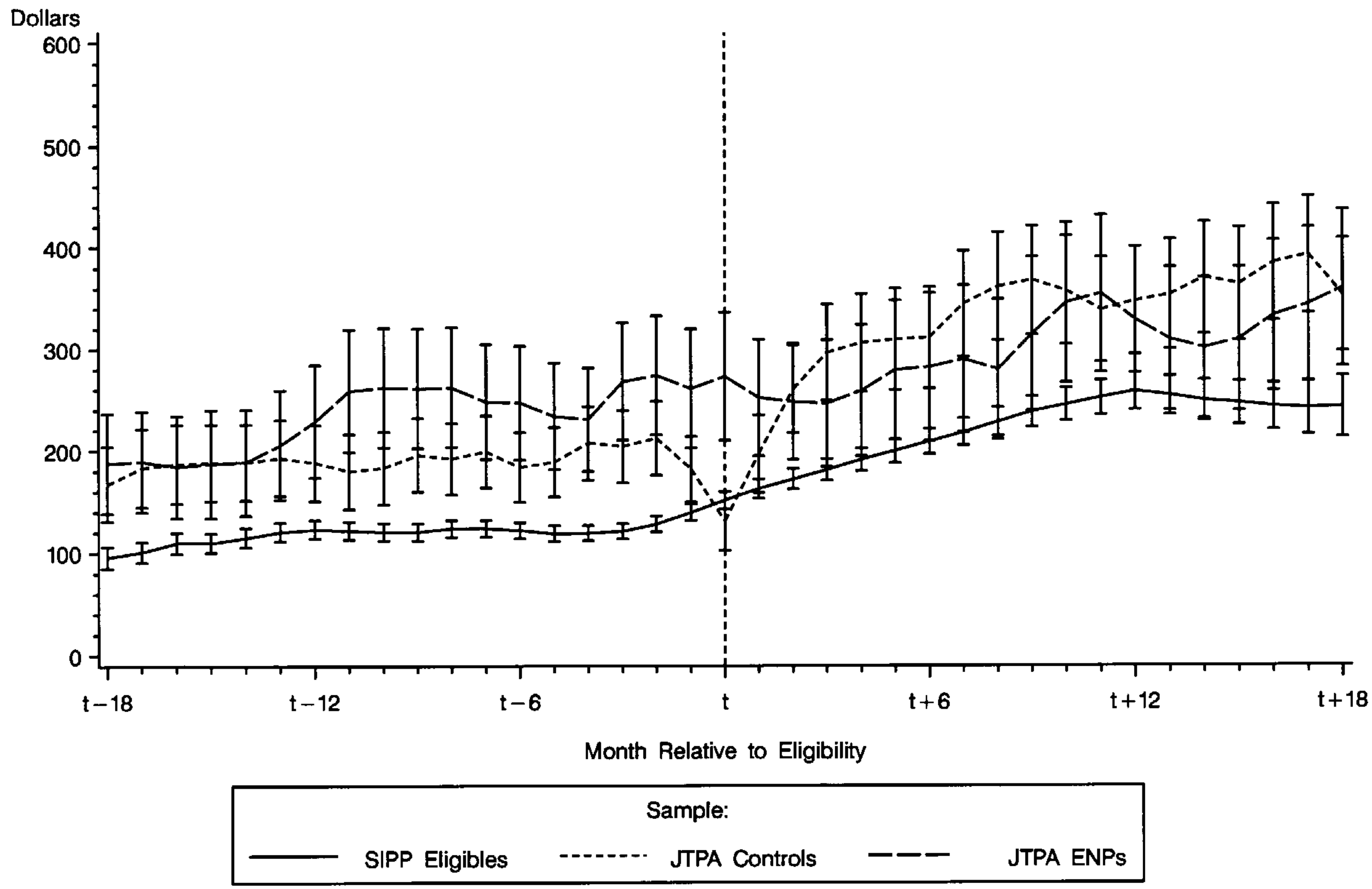
1. SIPP uses all JTPA-eligible person-month observations of respondents present in both the first and last months of the panel.
2. Controls are randomized-out participants from the National JTPA Study. Observations based on quasi-rectangular sample.
3. ENPs are JTPA-eligible non-participants at the same sites as the controls from the National JTPA Study. Observations based on quasi-rectangular sample.
4. Standard error bars +/- 2 standard errors of the means.
5. Top 1% of earnings trimmed in each month for ENP, control and SIPP samples.

FIGURE 1D

MEAN SELF-REPORTED MONTHLY EARNINGS

SIPP Eligibles and JTPA Controls and ENPs

Female Youths



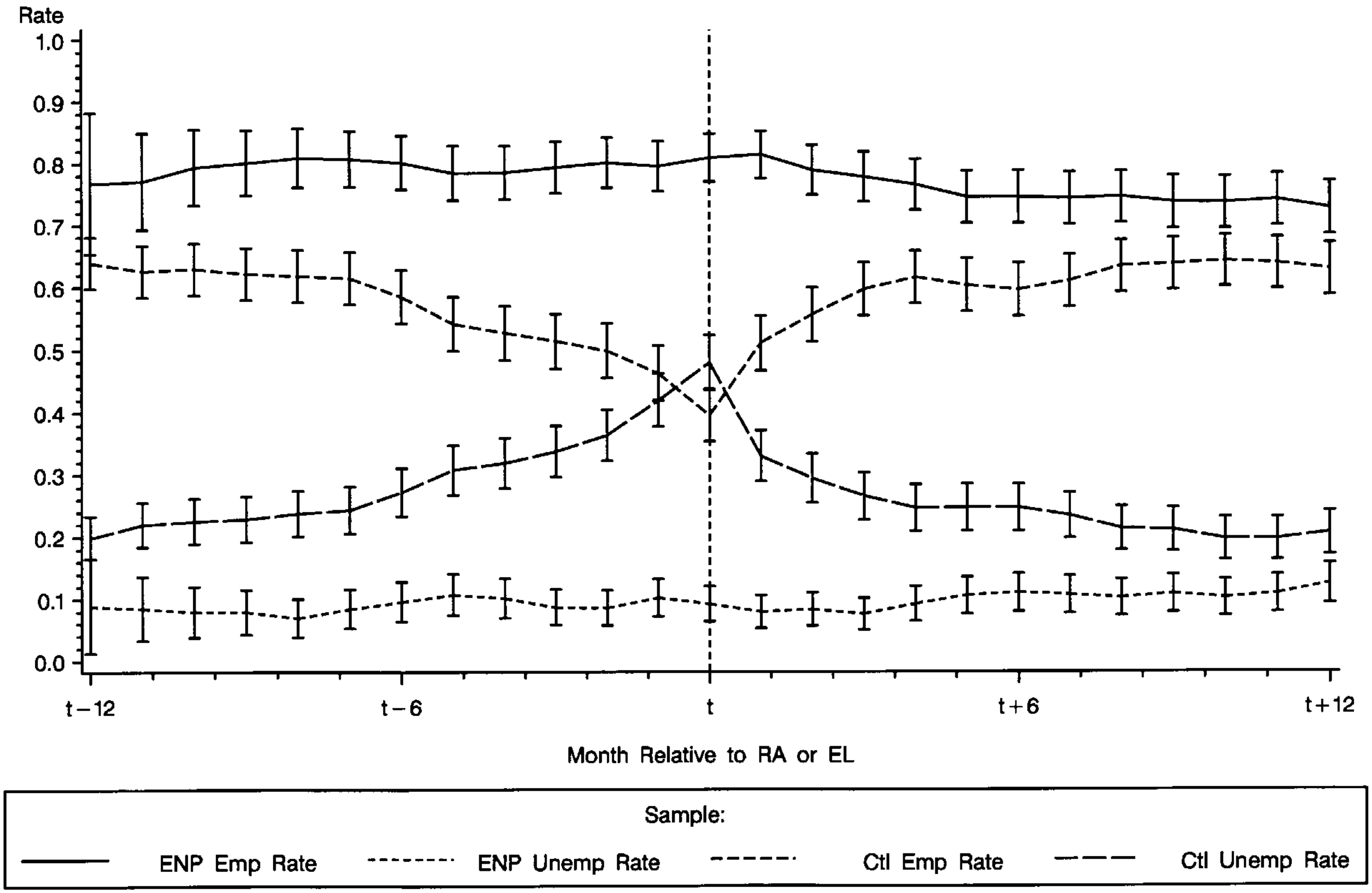
1. SIPP uses all JTPA-eligible person-month observations of respondents present in both the first and last months of the panel.
2. Controls are randomized-out participants from the National JTPA Study. Observations based on quasi-rectangular sample.
3. ENPs are JTPA-eligible non-participants at the same sites as the controls from the National JTPA Study. Observations based on quasi-rectangular sample.
4. Standard error bars +/- 2 standard errors of the means.
5. Top 1% of earnings trimmed in each month for ENP, control and SIPP samples.

FIGURE 2A

MONTHLY EMPLOYMENT AND UNEMPLOYMENT RATES

JTPA Controls and ENPs

Adult Males



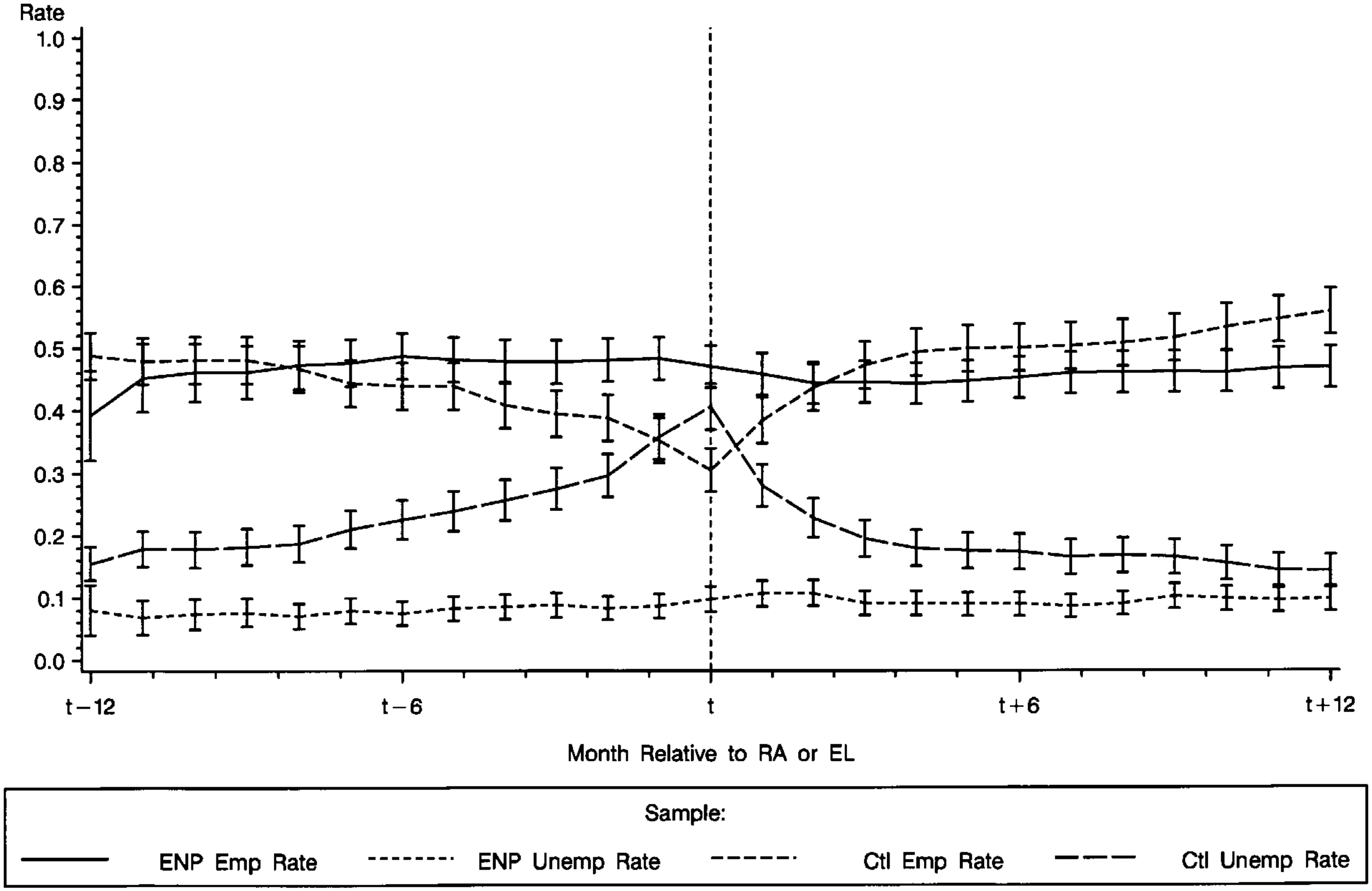
1. Controls are randomized−out participants from the National JTPA Study. Observations based on quasi−rectangular sample.
2. ENPs are JTPA−eligible non−participants at the same sites as the controls from the National JTPA Study. Observations based on quasi−rectangular sample.
3. Standard error bars +/− 2 standard errors of the means.

FIGURE 2B

MONTHLY EMPLOYMENT AND UNEMPLOYMENT RATES

JTPA Controls and ENPs

Adult Females



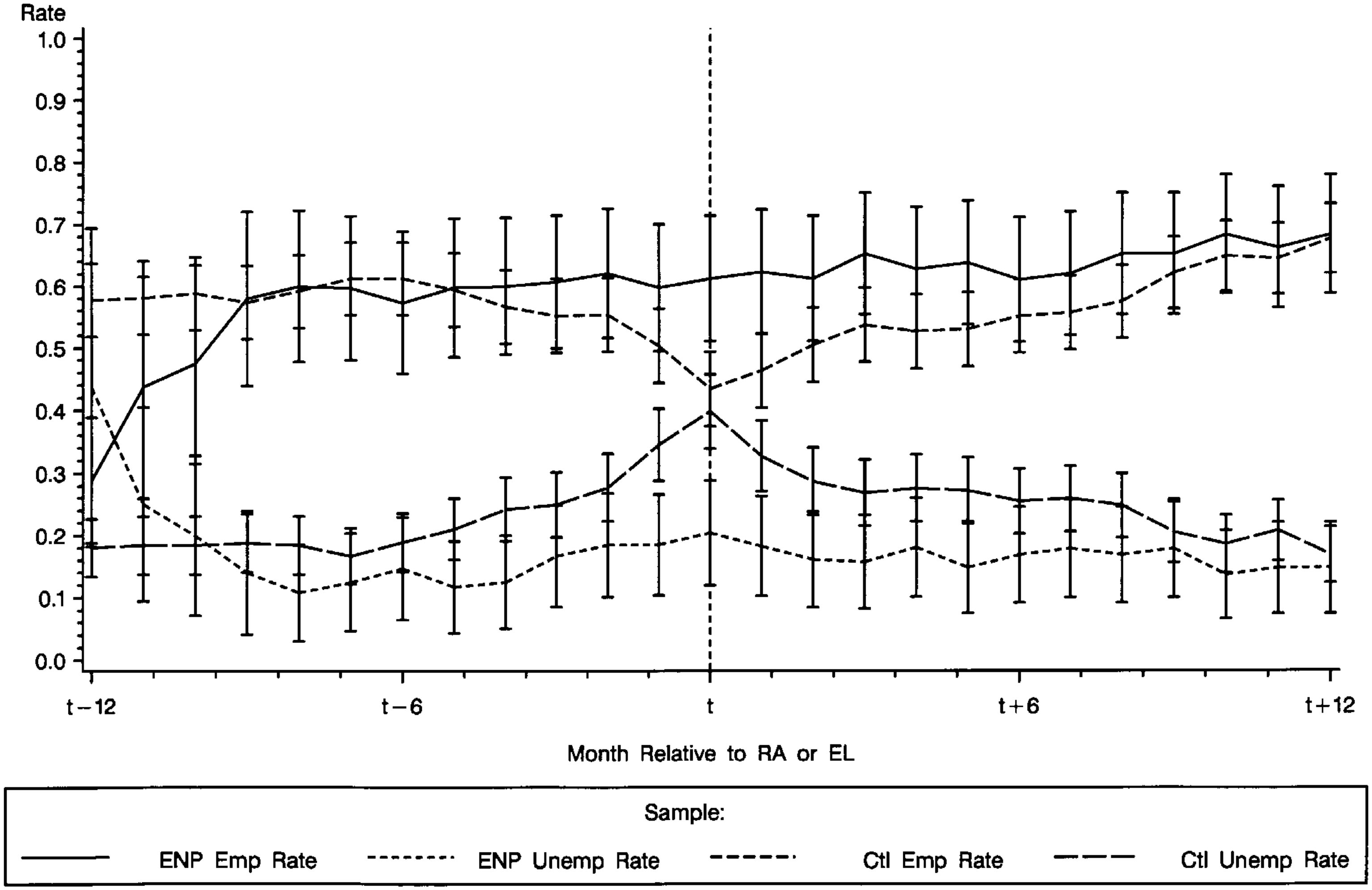
1. Controls are randomized−out participants from the National JTPA Study. Observations based on quasi−rectangular sample.
2. ENPs are JTPA−eligible non−participants at the same sites as the controls from the National JTPA Study. Observations based on quasi−rectangular sample.
3. Standard error bars +/− 2 standard errors of the means.

FIGURE 2C

MONTHLY EMPLOYMENT AND UNEMPLOYMENT RATES

JTPA Controls and ENPs

Male Youth



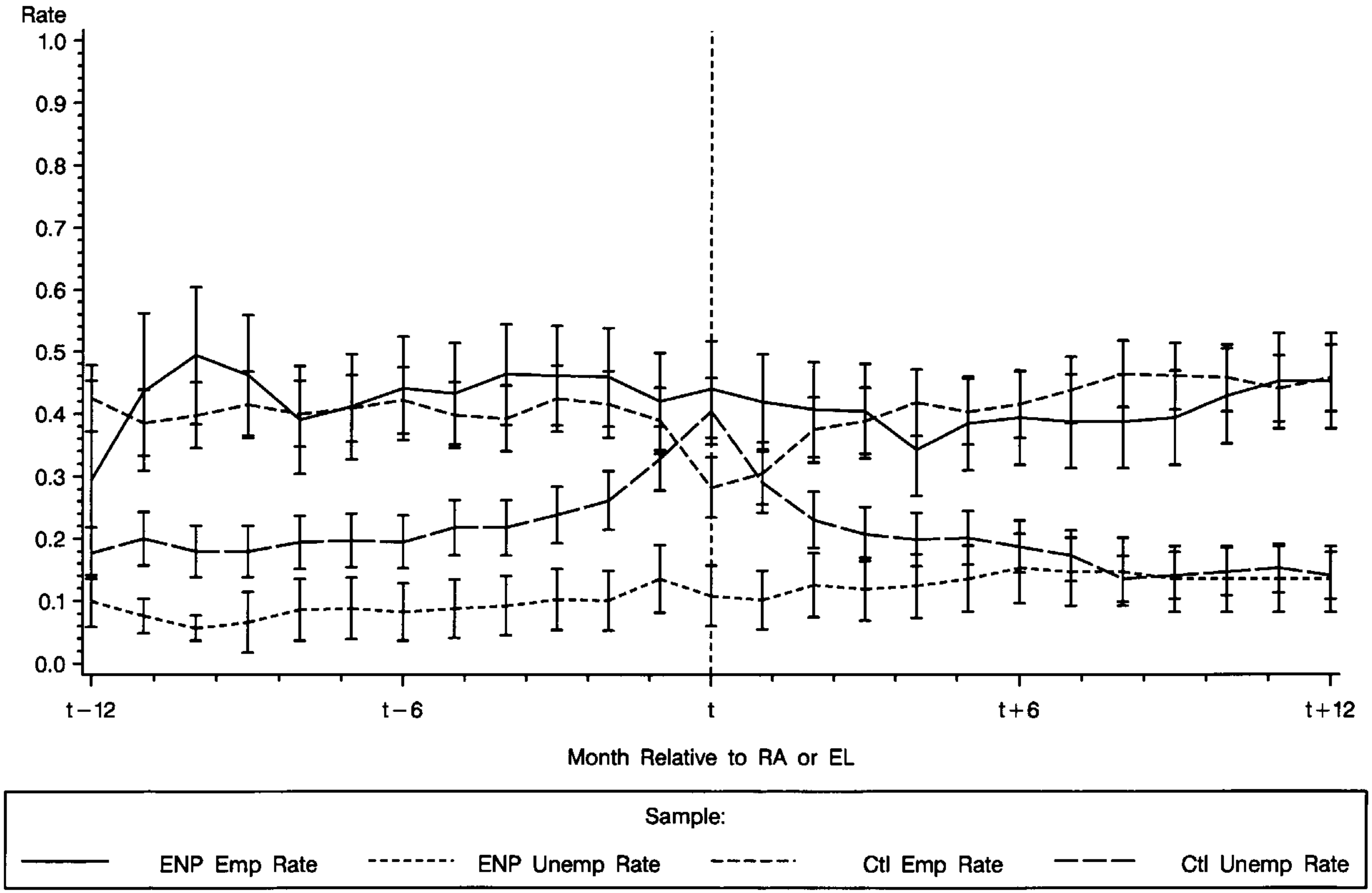
1. Controls are randomized-out participants from the National JTPA Study. Observations based on quasi-rectangular sample.
2. ENPs are JTPA-eligible non-participants at the same sites as the controls from the National JTPA Study. Observations based on quasi-rectangular sample.
3. Standard error bars +/− 2 standard errors of the means.

FIGURE 2D

MONTHLY EMPLOYMENT AND UNEMPLOYMENT RATES

JTPA Controls and ENPs

Female Youth



1. Controls are randomized−out participants from the National JTPA Study. Observations based on quasi−rectangular sample.
2. ENPs are JTPA−eligible non−participants at the same sites as the controls from the National JTPA Study. Observations based on quasi−rectangular sample.
3. Standard error bars +/− 2 standard errors of the means.

APPENDIX

DATA DESCRIPTION

1. SIPP Sample of JTPA Eligibles

We draw our national sample of persons eligible for JTPA from the 1986 Panel of the Survey of Income and Program Participation (SIPP). The SIPP is a continuing longitudinal self-weighting survey of the non-institutional population of the United States with a focus on current income and participation in social programs. The 1986 panel covers the period from October 1985 to March 1988.

We use eligibility Definition B in Devine and Heckman (1996), which captures only eligibility via economic disadvantage. We establish the eligibility status of each person in each month after the seventh month of the panel for which data are available. Eligibility cannot be established with certainty during the first six months of the panel because the requisite six months of prior data on family income are not available. To match the ENP sample, we exclude persons outside the 16 to 54 age range and those enrolled in junior high or high school. The graphs in Figures 1A to 1D and the estimates in Table 1 use a rectangular sample consisting of all eligible person-months of persons present in both the first and last months of the panel. We exclude observations with earnings imputed by the Census Bureau, as these imputations appear to be unreliable.

2. Eligible Non-Participant (ENP) Sample

The ENP sample is based on a sample of dwelling units drawn from the areas served by four of the sixteen training centers in the JTPA experiment: Corpus Christi, TX, Fort Wayne, IN, Jersey City, NJ and Providence, RI. At each center, the sampling frame excluded low poverty areas containing up to, but not more than, five percent of those with incomes at or below 125 percent of the poverty level in 1980. In the remaining areas served by each center, dwelling units were selected at random.

Attempts were made to collect data on all JTPA-eligible persons in the sampled dwelling units who were (1) eligible for JTPA via economic disadvantage, (2) 16 to 54 years of age, (3) not in junior high or high school and (4) not permanently disabled. Persons in the resulting sample had months of measured eligibility between January 1988 and December 1989. Only ENPs with valid earnings values for the 18th month before and the 18th month after measured eligibility were used in Figures 1A to 1D and for the difference-in-differences estimates in Table 1. The slightly different rectangular sample defined in Heckman, Ichimura, Smith and Todd (1998) is used for Figure 2. All ENPs with valid values for the relevant variables were used for the participation rates in Table 2 and the logit estimates in Tables 5 and 6.

3. Experimental Treatment and Control Group Samples

The experimental treatment and control group samples consist of persons randomly assigned at the four training centers in the JTPA experiment at which the ENP sample was drawn. Control group members were excluded from JTPA services for 18 months after random assignment. At the Corpus Christi and Fort Wayne centers, random assignment began in December 1987 and concluded in January 1989, while in Jersey City and Providence it ran from November 1987 to September 1989.

Controls with valid values of monthly earnings for the 18th month before and the 18th month after random assignment, and treatment group members with valid values for the 18th month after random assignment (no pre-random assignment data were collected for them), were used for Figures 1A to 1D and for the estimates in Table 1. The slightly different rectangular sample defined in Heckman, Ichimura, Smith and Todd (1998) is used for Figure 2. All controls with valid values for the relevant variables were used for the participation rates in Table 2 and the logit estimates in Tables 5 and 6.

4. Imputations

Missing values due to item non-response were imputed for the variables included in the estimation of the JTPA participation equations in Tables 5 and 6. Missing values of dichotomous variables were replaced with the *predicted probabilities* estimated in a logit equation. Missing values of indicator variables corresponding to categorical variables with more than two categories were replaced by the *predicted probabilities* obtained from a multinomial logit model. The models used to produce the imputations included indicators for race/ethnicity, age categories, receipt of a high school diploma or GED and training center, as well as interactions between control status and these variables. These variables were chosen because they had no (or very few) missing values in the sample. Imputed values were constructed separately for the four demographic groups.