

1996

# Three Essays on the Economics of Evaluating Social Programs: Dissertation Summary

Jeffrey Smith  
*University of Chicago*

---

# Three Essays on the Economics of Evaluating Social Programs

Jeffrey Smith

This dissertation consists of three essays on the evaluation of social programs. All three essays consider general evaluation questions in the specific context of evaluating the impact of government job training programs on the earnings of those who participate in them.

The first essay, jointly authored with James Heckman, examines three issues related to “Ashenfelter's dip”—the empirical regularity that the mean earnings of participants in employment and training programs generally decline during the period just prior to participation. This pattern was originally identified in Ashenfelter (1978) and has since been observed for participants in many other employment and training programs.

The first implication of Ashenfelter's dip for econometric evaluation research is the question it raises about what would have happened to participants had they not participated. The fundamental evaluation problem is that no person is ever observed simultaneously as both a participant and a nonparticipant in the program being evaluated. That is, in the context of a training program, the most that is ever observed is either what happens to the person if he or she did take training or what happens to the person if he or she did not take training, but never both. The difficult part of evaluations is constructing the unobserved counterfactual outcome that participants would have obtained had they not participated. This counterfactual is needed in order to determine the impact that the program has on its participants. Ashenfelter's dip makes it clear that participants are systematically different from nonparticipants in the period prior to participation, and raises the question of whether the earnings and employment losses reflected in the dip are permanent or transitory.

I address this question in Chapter II. Using experimental data from the recent National JTPA Study (NJS), I show what the counterfactual mean outcome is for participants. In a properly designed experiment, the outcomes of the experimental control group indicate what would have happened to participants had they not participated. Using the controls from the NJS, I show that for adult males and females, and for male and

female youth, the dip in mean earnings is transitory. In each case, the mean returns to its pre-dip level within six months after random assignment. For all the groups other than adult males, the dip is followed by growth in the earnings mean above pre-random-assignment levels in the post-random-assignment period.

This counterfactual earnings behavior has important implications for the most commonly used simple estimators of program impacts. These implications constitute the second issue addressed in Chapter II. I show that the pattern of earnings displayed by the control group indicates that before-after estimators, in which the pre-program experience of participants themselves serves as the estimate of the counterfactual outcome, are upwardly biased, with the extent of the upward bias depending on the particular “before” and “after” periods used to construct the estimates. This strong upward bias is consistent with the large positive impact estimates obtained in early evaluations of federal employment and training programs in the United States that used such comparisons.

More recent evaluations use a comparison group of nonparticipants whose earnings behavior serves as a benchmark against which to compare the earnings behavior of the program participants. The simplest, and most widely used, comparison group estimator is the “difference-in-differences” estimator, in which the before-after earnings change of the comparison group is subtracted from the before-after earnings change of the participants. This estimator is motivated by a model in which persons select into a program based on a fixed, person-specific component of earnings. This component is then differenced out in the estimation procedure. I evaluate the performance of the difference-in-differences estimator relative to the experimental estimates using two different comparison groups. The first consists of JTPA-eligible nonparticipants (ENPs) from four of the sites in the NJS and the second consists of a national sample of

---

Jeffrey Smith received his Ph.D. from the University of Chicago. He is an Assistant Professor in the Department of Economics at the University of Western Ontario. Mr. Smith's dissertation advisor was James J. Heckman.

persons eligible for JTPA drawn from the 1986 Full Panel of the Survey of Income and Program Participation (SIPP). Both samples are superior to the comparison group data used in earlier studies. In particular, the ENPs are drawn from the same local labor markets and administered the same surveys as the experimental sample. Both the ENPs and the SIPP eligibles are all known to be eligible for JTPA. Though the ENP sample generally performs better than the SIPP sample in the sense that the estimates obtained using it are closer to the experimental benchmark, for both samples the difference-in-differences estimates differ substantially from the experimental estimates. These differences result primarily from the post-program earnings growth observed for the controls. This growth is not observed in the comparison group samples for most demographic groups, with the result that the difference-in-differences estimates also tend to show an upward bias. Furthermore, the estimates are quite sensitive to the "before" and "after" periods used in constructing them. This sensitivity results in part from the post-program earnings growth, and in part from the effects of Ashenfelter's dip, which is also not observed in either comparison group.

The regular appearance of Ashenfelter's dip among participants in a wide variety of employment and training programs has led later researchers such as Ashenfelter and Card (1985) and Card and Sullivan (1988) to focus on employment and earnings processes as the driving forces behind participation in employment and training programs when constructing econometric models of the participation process. The final section of Chapter II examines what can be learned about the determinants of participation in JTPA from the ENP and control data from the NJS for the four training centers for which both samples are available.

This analysis reveals that the earlier literature's focus exclusively on earnings and employment, which was motivated in part by Ashenfelter's dip and in part by the limitations of the available data, leaves out an important part of the story. I find that labor force status, defined as the usual CPS trinity of employed, unemployed, and out of the labor force, plays an important role in the participation process beyond that played by earnings or employment. Using as a metric the ability to predict who among the combined ENP and control samples will and will not participate, I find that labor force status patterns in the seven months up to and including the month of the participation decision do better than measures based solely on earnings or employment, particularly for groups other than adult males. The patterns most likely to lead to participation in JTPA are a recent transition into unemployment from either

employment or from out of the labor force. Among those employed or out of the labor force at the time of the participation decision, those who have recently entered these states are relatively more likely to participate in JTPA than those who have not. Thus, participants tend to be those whose labor force status is in flux, particularly the recently unemployed. This finding is consistent with the fact that both JTPA and its predecessor programs CETA and MDTA provide not only traditional classroom training, but also job search assistance and placement services that would be of interest to unemployed persons looking for immediate employment rather than for more traditional classroom training.

Chapter III, which is also jointly authored with James Heckman, examines the determinants of selection into the JTPA program more broadly. Rather than focusing on a single transition from eligibility to acceptance as in the first essay, the enrollment process is decomposed into a series of stages, from eligibility for JTPA, to awareness of JTPA, to acceptance into JTPA and finally to formal enrollment in the program. Decomposing the process in this way indicates the sources of observed demographic differences in JTPA participation rates that have troubled some observers of the program. It also sheds light on the extent to which demographic differences result from individual self-selection or from the actions of program administrators. Because JTPA is not an entitlement program like AFDC or Food Stamps, program administrators have substantial discretion over whom to serve and how to serve them. There is a concern that the structure of the bureaucratic performance standards system within JTPA encourages program bureaucrats to "cream-skim" by bringing in only the most employable persons within the JTPA-eligible population.

Chapter III begins with a systematic analysis of the determinants of the transition at each stage of the overall process of enrollment in JTPA, while the final section presents decompositions that combine some or all of the stages and reveal the relative importance of particular factors at different stages. This analysis yields several important findings. First, I find some evidence consistent with cream-skimming by program bureaucrats. Most of this evidence is concentrated at the stage from acceptance into the program (indicated here by random assignment) and formal enrollment. Second, I find that informational barriers such as lack of fluency in English and low levels of completed schooling act to discourage participation conditional on eligibility, but that these differences do not fully account for the differentially low rate of Hispanic enrollment in JTPA found in other studies. Third, I show that the importance of labor force status

transitions, already pointed out in Chapter II, is concentrated at the stage from awareness of JTPA to acceptance into it. Fourth, I show that the participation patterns of AFDC and Food Stamp recipients differ markedly from those of other groups. They are relatively more likely to be aware of the program than other eligibles, but less likely to enroll in the program conditional on acceptance. Finally, and more broadly, this analysis shows that manipulation of program eligibility rules constitutes a weak tool for increasing the participation of particular groups. Some groups, including some cited as being particularly in need of JTPA services, have much lower rates of participation conditional on eligibility than do others. Increasing the participation of these groups requires more active measures than simply broadening the eligibility rules to include them.

In Chapter IV, I focus on the measurement of earnings dynamics among the low-income population eligible for training in JTPA. This essay is part of a larger project that also examines the measurement of earnings levels within this population. Accurate measurement of both the level and temporal pattern of earnings is crucial to obtaining reliable estimates of program impact and of the determinants of program participation.

Chapter IV compares the dynamics of mean earnings in two samples of persons eligible for training under JTPA. The first sample is drawn from the 1988 Full Panel of the Survey of Income and Program Participation (SIPP), while the second is a sample of eligible nonparticipants (ENPs) at four of the sixteen sites in the National JTPA Study. These are the same two samples used as comparison groups in Chapter II—with the exception that the SIPP sample is drawn from the 1988 panel rather than the 1986 panel. In that chapter, I show that the structure of the eligibility rules for the JTPA program, which require either low family income in the six months prior to application or participation in certain means-tested transfer programs at the time of application, can lead to a dip in the mean family income of eligibles in the months just prior to the month of measured eligibility. Examining this result empirically in Chapter IV, I find that this dip appears in the mean individual earnings of adult male and adult female SIPP eligibles but not in the mean individual earnings of the ENPs. The failure of the dip to appear in the data on the individual earnings of youth can be explained by the fact that the earnings of youth typically represent only a small portion of total family income, which is the income that counts for JTPA eligibility. In contrast, the absence of a dip in the mean earnings of adult ENPs is more difficult to account for.

I examine two alternative explanations for the difference in earnings dynamics between the two samples. The first is that the lack of a dip in the mean earnings of adult ENPs is accounted for by the exclusion of areas with low poverty rates from the ENP sampling frame. Persons eligible for JTPA in areas with low poverty rates are differentially nonpoor. At the same time, in the SIPP sample, nonpoor eligibles show a much more powerful dip in mean earnings prior to measured eligibility, as they must in order to become eligible for the program. I show that the number of nonpoor eligibles excluded at the four sites in the National JTPA Study is too small to account for the absence of a dip in mean earnings in the ENP sample.

The second explanation builds on differences in the survey instruments used to collect earnings data on the two samples. The survey administered to the ENPs resembles that used for the National Longitudinal Survey of Youth in that both surveys collect earnings information indirectly through questions about particular job spells. On the two ENP surveys, for each job reported in an interview, one value for average hours worked per week is collected. Two values for the rate of pay, the starting wage and the current or ending wage, are collected for most jobs paid by the hour. For other jobs, only the pay period (week, month, etc.) and the usual pay per period are collected. Collecting the data in this way implicitly removes all variation within job spells in hours worked or in rate of pay. The additional information collected in both surveys on overtime pay, tips and bonuses, as well as the limited information on weeks worked without pay collected on the follow-up survey, also has no variation within job spells. Thus, the structure of the survey instruments administered to the ENPs forces all of the temporal variation in measured earnings for individual ENPs to result from either job loss, job gain, or crossing the seam between the time periods covered in the two surveys.

In contrast, the SIPP survey collects earnings information directly from questions about earnings in each month on each of the two jobs for which the respondent reports working the most hours in each four-month SIPP survey reference period. Information on earnings from businesses is collected in the same way. This method of collecting earnings information allows variation in earnings from month to month on each job. Furthermore, by having interviews every four months, the SIPP respondents are less likely to forget periods of increased or reduced earnings that might be lost in the ENP surveys due to the long (five years for the baseline survey and 18 months or more for the follow-up survey) recall periods required.

If the dip in mean earnings prior to eligibility results primarily from job loss and job gain, then both surveys should, in principle, pick it up. On the other hand, if the dip results primarily from within-spell reductions in earnings due to periods of absence from work without pay, reduced overtime or temporary layoff, then only the SIPP instrument will capture it in the data. To shed some light on which of these two hypotheses is the correct one, I smooth out the SIPP data by taking the total earnings from each employment spell and assigning an equal fraction to each month of the spell. This replicates the smoothing induced in the ENP earnings data by the ENP survey instrument. Examination of the smoothed SIPP data reveals that the smoothing process removes the dip in mean earnings in the months prior to measured eligibility. The same result obtains when the SIPP earnings are smoothed over job spells rather than employment spells. These findings constitute strong evidence that the lack of a dip among the ENPs results from the failure of the survey instruments administered to that sample to effectively

capture the within-spell changes in earnings that underlie the dip in the mean earnings of JTPA eligibles. More generally, these findings show that choices about survey design have important implications for our ability to accurately measure earnings dynamics among the poor.

### References

- Ashenfelter, Orley. 1978. "Estimating the Effect of Training Programs on Earnings." *Review of Economics and Statistics*, 60, 47-50.
- Ashenfelter, Orley, and David Card. 1985. "Using the Longitudinal Structure of Earnings to Estimate the Effect of Training Programs." *Review of Economics and Statistics*, 67, 648-660.
- Card, David, and Daniel Sullivan. 1988. "Measuring the Effect of Subsidized Training Programs on Movements In and Out of Employment." *Econometrica*, 56(3), 497-530.