

Differential Effects of Welfare to Work Programs: Identification with Unknown Treatment Status*

Oscar A. Mitnik[†]- University of Miami

December 15, 2005

Abstract

This paper estimates the treatment effects of the two different training strategies followed by “Welfare-to-Work” (WTW) programs: Labor Force Attachment (LFA) and Human Capital Development (HCD). A key problem in estimating the effects of these two types of programs is that the available data sources do not identify which individuals have been subject to training. This paper presents a non-experimental econometric methodology that allows identifying the treatment effects even under unknown treatment status.

The results, from estimations based on administrative data for California, suggest that the LFA programs have short term positive effects that fade out around two years after entry into welfare, while the effects of HCD programs are non-significant or negative, and the effects of being on welfare without receiving any training appear as negative in the first two years after welfare entry, and become positive afterwards. After welfare reform, however, the effects of both training programs appear as non significant, with the effect of no training strongly positive. This might be a consequence of the interaction of the new environment after welfare reform with the WTW programs.

Keywords: Training Programs; Welfare Programs

JEL Codes: I38, J24

*I am grateful to V. Joseph Hotz and Guido Imbens for their encouragement and support, and for their invaluable guidance. I also would like to thank Sandra Black, Moshe Buchinsky, Janet Currie, Eduardo Fajnzylber, Jeffrey Grogger, Enrico Moretti, Charles Mullin, Juan Pantano, and participants in the Applied Microeconomics Workshop at UCLA for their comments and suggestions. All remaining errors are the author’s responsibility. Financial support from the UC Institute for Labor and Employment is gratefully acknowledged.

Part of the data in this study was used with the permission of the California Department of Social Services. The opinions and conclusions expressed herein are solely those of the author and should not be considered as representing the policy of any agency of the California State Government

[†]Department of Economics, University of Miami, omitnik@miami.edu, <http://moya.bus.miami.edu/~omitnik>

1 Introduction

One of the most prominent features of the 1996 reform of the welfare system in the U.S. was the emphasis on employment as the key to self-sufficiency. A key component in the reform was the “Welfare-to-Work” (WTW) programs, which provide welfare recipients different types of training that are supposed to ease the transition to the labor force and eliminate the need for aid use. The type of services offered by the WTW programs follow two different approaches that have potentially different implications on wage dynamics and skill formation. The Labor Force Attachment (LFA or work-first) type of programs try to increase the insertion of welfare recipients into the labor force by providing job search training and assistance, while the Human Capital Development (HCD) type of programs are oriented to increase the trainees human capital by offering longer duration basic skills and vocational training programs.

In this paper I evaluate the effects of WTW programs once broadly implemented, estimating the differential effect of the LFA and HCD training strategies. This allows to assess whether these programs help welfare recipients, and if so by how much. In addition, part of the period analyzed is after the implementation of welfare reform allowing to compare pre and post welfare reform effects. A key problem in estimating the effects of these two types of programs is that the available data sources do not identify which individuals have been subject to training. I solve this by resorting to a non-experimental econometric methodology that allows identifying the treatment effects even under unknown treatment status.

The question of the relative effects of LFA and HCD training programs has been of interest for a long time, and has had important consequences. Based on the results of a series of experimental studies on the effects of training on welfare recipients performed in the early 1990s by the Manpower Demonstration Research Corporation (MDRC), welfare reform had a bias towards Labor Force Attachment type of programs. One of the first, and very influential, studies was on California’s Greater Avenues for Independence (GAIN) Program, which showed very strong positive effects on the employment and earnings, and a decrease in aid reliance, of the individuals trained in comparison to a randomly assigned control group. These effects were particularly strong in

Riverside County, which favored a strong work-first approach. Although the success of Riverside's work-first approach had an influential effect on the design of welfare reform, the GAIN experiment was not designed to compare the effectiveness of the two training approaches. Subsequently, MDRC conducted a multi-site evaluation, the National Evaluation of Welfare-to-Work Strategies (NEWWS), with individuals participating in the JOBS Program,¹ randomized between 1991 and 1994. Three sites of the NEWWS evaluation were chosen to explicitly study the differential effects of LFA versus HCD programs by randomly assigning individuals to one of the two treatments or to a control group.² The results from this evaluation show that LFA programs have better outcomes two years after training (approximately a 10% increase in employment rates), but after five years the differences between LFA and HCD programs disappear (with very modest effects overall). This pattern is also identified in a recent survey by Grogger, Klerman and Karoly (2002) on the effects of mandatory work-related activities for welfare recipients (based mostly on experimental studies).³

However, the existing evidence is based on relatively small experiments in a few sites, which raises the question of whether these results would change when evaluating a program applied at full scale. Also, welfare reform introduced significant changes like time limits and financial incentives to work, which could make the results from the previous demonstration programs not applicable to the post-welfare reform period.⁴ In particular, the interaction of the new rules after welfare reform and the work and training requirements of the WTW programs may have effects not captured by these evaluations.

The question of how much can be learned from the results of experimental programs in terms of the broad implementation of similar programs to other locations, to individuals with different characteristics and for different economic environments has received considerable attention recently in the literature. Dehejia (2003) analyzes the GAIN experiment and concludes that Riverside's treatment effects are very difficult to predict outside Riverside, implying that the idiosyncratic characteristics of the treatment in Riverside might be difficult to replicate elsewhere. The effect of the heterogeneity in treatments in hampering the extrapolation of experimental studies is also analyzed by Hotz, Imbens and Mortimer (1999) and Hotz, Imbens and Klerman (2000), who compare

the control groups in different sites where experiments were conducted to account for observed individual differences. In both studies the authors find important effects of treatment heterogeneity, after controlling for population differences.^{5,6}

In this paper I utilize to a non-experimental strategy that actually exploits the heterogeneity of treatments received by welfare recipients not only across sites but also across periods. I extend a methodology proposed by Heckman and Robb (1985) for single treatments, to the case of multiple treatments and apply it to welfare recipients in California. The period analyzed includes both the GAIN and WTW programs (before and after welfare reform).

One reason why there are no studies on the effectiveness of WTW programs, other than those based on demonstration programs, is that administrative and survey data do not normally identify which individuals are subject to training. The econometric methodology in this paper allows for identifying the treatment effects of the two types of training programs of interest, even while not knowing the identity of the trainees. This is accomplished by using two different administrative data sources, one that provides individual level information on welfare use and earnings in California, and another that provides information on the proportions of individuals that received each type of training, by county and time period. The temporal variation in the probability of receiving training is exploited to identify the differential training effects.

It seems important to try to use non-experimental methods to overcome the data restrictions (namely the unavailability of information on who are the treated individuals), for several reasons. First, after welfare reform is less likely that any experimental evaluation will be possible, given that experimental evaluations are very expensive to implement for a full scale program. Second, after welfare reform welfare recipients face time limits, which would make it unethical to ban services to certain welfare recipients. Third, except for the case of NEWWS in which three sites were chosen to conduct a random experiment comparing the two training strategies, other experimental evaluations were not designed to answer the specific question of this paper, which means that even using experimental data, non-experimental methods would be necessary to obtain an answer.⁷

From a policy point of view this research provides valuable information on the effects of different

WTW strategies, and on the likelihood that welfare recipients attain self-sufficiency, TANF's stated main objective. The importance of the answer to this question is highlighted by the renewed emphasis on work requirements in the political discussions on the reauthorization of the TANF Program.

As a first step in estimating the differential treatment effects of the training strategies, I estimate the effects of first time entry to welfare, for different entry cohorts, using propensity score weighting estimators for comparing individuals treated (welfare entrants) to control individuals (those who have never been on welfare up to this point) of similar characteristics. This first step is interesting in itself, showing that over time it has decreased (from 9 to less than 4 quarters) the amount of time required for welfare recipients to recover from the employment and earnings shocks that determine their entry to welfare in the first place. This recovery period is even faster after welfare reform. In a second step these entry to welfare effects are explained by the heterogeneity in treatments across entry cohorts and counties, to obtain the differential treatment effect of the training programs. The results show that the LFA programs have short term positive effects that fade out around two years after entry into welfare, while the effects of HCD programs are non-significant or negative, and the effects of being on welfare without receiving any training appear as negative in the first two years after welfare entry, and become positive afterwards. After welfare reform, however, the effects of both training programs appear as non significant, with the effect of no training strongly positive. This might be a consequence of the interaction of the new environment after welfare reform with the WTW programs, and highlights that more is needed to be done to better understand these interactions.

This paper is organized as follows. The next section gives some background information on the programs evaluated. The third section discusses the identification strategy and the fourth section describes the data used. The fifth section presents the estimation strategy and the sixth section details some tests of the identifying assumptions. Finally, the seventh section analyzes the estimation results, and the eighth section presents the conclusions.

2 The GAIN and WTW programs

Training was offered to welfare recipients in California in the 1990s through two programs. California's version of the JOBS Program, under the Aid for Families with Dependent Children (AFDC) Program, was the Greater Avenues for Independence (GAIN) Program, in which training was the main component. It started in 1989 and it was succeeded in 1998 by the Welfare to Work (WTW) Program, as part of the California Work Opportunity and Responsibility to Kids (CalWORKs) program, California's version of the Temporary Assistance for Needy Families (TANF) Program.⁸ Both programs offered different types of services to welfare recipients who were mandated to participate (except parents of small children), or face financial sanctions. However, under GAIN counties faced severe funding constraints, and in some counties a big proportion of the caseload remained not served.

Under the rules of CalWORKs every adult is required to participate in the Welfare to Work Program, which implied that counties had to expand their programs to accommodate all the adult caseload.⁹ The activities that the programs offered included, among others, job search and job readiness assistance, on the job training and subsidized employment, vocational education training, adult basic education, English as a second language, and classes for preparing to take the General Education Diploma (GED) exam.

The training activities have been classified in two groups: work-oriented, termed Labor Force Attachment (LFA), and education-oriented, termed Human Capital Development (HCD), types of training. Typically LFA training has a shorter duration and is less expensive to provide than HCD training (see Hamilton, Freedman et. al., 2001 for a discussion of both approaches).

Counties in California were given a great degree of freedom over the design of their GAIN and WTW Programs. This caused a remarkable variation both across counties and across time in the proportion of the adult caseload that participated in any activity, and that participated in LFA and HCD types of training. The variation from 1995 through 1999 Q2 in the proportion of adults receiving any training, and the two types of training is presented in Figure 1.¹⁰ As can be seen in the figure, there has been an increasing trend in the proportion of individuals trained in general,

but more marked for LFA training. When the CalWORKs program started (January 1998), there was a larger jump in absolute terms for LFA than for HCD training, although the latter increased more in relative terms.¹¹

Note that the figure presents data for all the adults on welfare in a particular period of time. Unfortunately there is no data available that breaks these proportions between new entrants and non-entrants to welfare. The identification strategy and empirical analysis will be based, however, in comparing cohorts of new entrants to welfare, because is the only way to assure that the effects of potential prior treatments are not confounded as part of the treatment effects calculated. Moreover, both the GAIN and WTW programs (with more emphasis the second one) were devised such that each new entrant to welfare was assessed and assigned to a treatment in a relatively short period of time. However, under the GAIN program, counties in which the available resources did not permit the treatment of a large percentage of the caseload had a long waiting list for welfare recipients assigned to a particular training. This is reflected in large proportions for the “no training” category, which will be explicitly considered in the next section.¹²

3 Identification strategy

The objective of this paper is to compare the treatment effects of the two approaches followed by the training programs for welfare recipients: Labor Force Attachment (LFA) and Human Capital Development (HCD). Because there is no available data that identifies nor which individuals received training, nor the type of training received, in this paper I will use administrative datasets that contains individual-level and county-level information to study the differential effects of these training strategies. The individual-level data includes demographic characteristics, welfare use and earnings of all welfare recipients in California, while the county-level data gives the percentage of welfare recipients that received each type of training. A key feature of the individual-level dataset is that it provides information on welfare recipients for relative long periods before their entry to welfare. This will allow me to use future entrant cohorts as controls for present entrants.

As with any non-experimental evaluation, the identification is based on assumptions that are in

most of the cases untestable (although some implications of these assumptions will be tested). There is an ongoing debate since the 1980s regarding the merits of non-experimental versus experimental evaluations launched by the influential study by Lalonde (1986).¹³ The problem under analysis is particular because only an experimental evaluation that assigns individuals randomly to the two types of activities would be free of non-experimental assumptions. Most prior evaluations of training programs for welfare recipients were not designed to give this answer, which implies that also untestable non-experimental assumptions would be needed to evaluate the differential effects of the programs, even using experimental data. The exceptions are the three sites in the NEWWS evaluation, which provided results that are non conclusive respect to what would be the effects under a fully implemented program. In particular, welfare reform generated several rules changes that can have important interaction effects with mandated work programs. In addition, after welfare reform welfare recipients face time limits, which would make it unethical to ban services to certain welfare recipients for the purpose of conducting an experiment.

3.1 Setup of the problem

Following the convention of the literature, everything will be expressed in terms of potential outcomes. $Y_{it}(0)$ will represent the potential outcome of individual i in time period t , if the individual has not received any treatment. The possible treatments are: to receive training in a LFA program, to receive training in a HCD program, or to be a welfare recipient, but to not receive any type of training. The potential outcomes associated to these treatments will be denoted by $Y_{it}(m)$, where $m = \{L, H, N\}$ represents the three possible treatments, respectively. Welfare reciprocity, without receiving training, is considered a treatment in this context, because welfare recipients receive other services anyway (primarily a cash grant and health insurance).

The typical evaluation problem is to estimate the effect of a treatment on individual i respect to the case where no treatment is received, $\theta_i(m) = Y_{it}(m) - Y_{it}(0)$. Because no individual can be observed in two states at the same time, only an average treatment effect can be actually calculated. In this particular case, the three treatments can be received only conditional on being a welfare

recipient, which means that the parameter of interest will be an Average Treatment on the Treated Effect (TT) (i.e. given welfare receipt) of treatment m

$$\begin{aligned}\theta(m) &= E[Y_{it}(m) - Y_{it}(0)|W_i = 1] \\ &= E[\theta_i(m)|W_i = 1]\end{aligned}\tag{1}$$

where W_i is a dummy equal to 1 if the individual entered welfare before time t , and equal to 0 otherwise. This parameter gives the expected effect of treatment m on a random individual, conditional on the individual being eligible to receive the treatment (i.e. the individual was a welfare recipient before t). This needs to be distinguished from what will be referred in this paper as a Conditional Average Treatment on the Treated Effect (CTT),

$$\begin{aligned}\theta^*(m) &= E[Y_{it}(m) - Y_{it}(0)|W_i = 1, T_i(m) = 1] \\ &= E[\theta_i(m)|W_i = 1, T_i(m) = 1],\end{aligned}\tag{2}$$

which gives the expected effect of treatment m conditional on the individuals that were subject to this treatment ($T_i(m)$ is a dummy variable equal to 1 if the individual received treatment m , and equal to 0 otherwise).

While $\theta^*(m)$ is an interesting parameter in itself to evaluate the effectiveness of a training program, it is not the right parameter to use when comparing the effectiveness of two treatments.¹⁴ To compare the effectiveness of, say, treatment m respect to treatment m' , it is necessary to define the Differential Treatment on the Treated Effect (DTT)

$$\begin{aligned}\gamma(m, m') &= E[Y_{it}(m) - Y_{it}(m')|W_i = 1] \\ &= E[\theta_i(m)|W_i = 1] - E[\theta_i(m')|W_i = 1] \\ &= \theta(m) - \theta(m'),\end{aligned}\tag{3}$$

which gives the differential effect of one treatment respect to the other, for an individual randomly selected among the individuals that were welfare recipients before t . In this paper the parameter of interest will be $\gamma(L, H)$, the differential effects of LFA versus HCD type of training programs.

To estimate the counterfactual outcome, i.e. what would have been the outcome of the individuals subject to a particular treatment if they had not received the treatment, it is necessary to make assumptions about the state not observed, using for example the average outcome observed for another time period and/or group of individuals. The problem in the case under analysis is that is not even possible to observe which individuals are subject to treatment, which means that the potential outcomes under treatment for the treated individuals is not identified. It is known that the training programs to be evaluated are only available to welfare recipients, which makes it possible to express the potential outcome at time t of an individual that entered welfare at some point before t , $Y_{it}(W)$, as

$$Y_{it}(W) = T_i(L)Y_{it}(L) + T_i(H)Y_{it}(H) + T_i(N)Y_{it}(N), \quad (4)$$

where by definition $T_i(N) = 1 - T_i(L) - T_i(H)$. Therefore, the observed outcome for an individual i at time t , Y_{it} can be expressed as

$$Y_{it} = (1 - W_i)Y_{it}(0) + W_iY_{it}(W). \quad (5)$$

This suggests a way of identifying the TT and DTT by first estimating the overall “treatment on the treated effect” of entering into welfare, and then exploiting the differences in the probability of receiving training by different cohorts of new entrants to welfare, across time or across counties.

To develop this strategy it is necessary to be more precise with the timing and location issues. All the variables (when adequate) will have superindexes c and t , that will identify the county where a persons first entered on welfare and the time period in which this person entered, respectively (for example Y_{it+j}^{ct} will indicate the observed outcome at period $t + j$ for the individual i , who entered for the first time on welfare in county c at time t).

A key assumption will be that training can only be received in the first period in which an individual goes into welfare. This “first period”, though, could be defined in a way flexible enough to be consistent with different time frames for the training activities.

The way in which identification can be attained will vary in some way depending on whether

the treatment effects are assumed to be homogeneous or heterogeneous. The homogeneous case will be analyzed first, and then the modifications that need to be introduced for the heterogeneous case.

3.2 Homogeneous treatment effects

If the treatment effects are homogeneous across individuals, as it is often considered in the literature, then the difference between Average Treatment on the Treated Effect (TT) and Conditional Average Treatment on the Treated Effect (CTT) disappear, $\theta(m) = \theta^*(m)$ as it is clear from (1) and (2). In this case the identification can be attained in two steps. In the first one, considering “welfare entry” as the treatment, the TT effect of welfare j periods after entering welfare, for the individuals that entered welfare in county c at time t , can be estimated as

$$\Delta_j^{ct} = E[Y_{it+j}^{ct}(W) - Y_{it+j}^{ct}(0) | W_i^{ct} = 1]. \quad (6)$$

This treatment effect, can be decomposed, using (4) as a weighted average of the TTs j periods after entering welfare of each of the three possible treatments received by welfare entrants at time t ,

$$\Delta_j^{ct} = P_L^{ct} \theta_j^{ct}(L) + P_H^{ct} \theta_j^{ct}(H) + P_N^{ct} \theta_j^{ct}(N), \quad (7)$$

where $P_m^{ct} \equiv \Pr(T_i^{ct}(m) = 1)$ is the probability that each entrant received treatment m in county c at time t , and $\sum P_m^{ct} = 1$.

The second step exploits the fact that the probabilities of receiving training present variation not only across counties, but also across time within counties, and is an extension to the multiple treatment case of the strategy proposed by Heckman and Robb (1985) for cases with a single treatment. Assuming that the TTs vary across counties but not over time, i.e. $\theta_j^{ct}(m) = \theta_j^{ct'}(m) = \theta_j^{ct''}(m) = \theta_j^c(m)$ for at least three cohorts of entrants at t , t' , and t'' , and given that the probabilities of receiving treatment are observed, then $\theta_j^c(m)$ can be identified, as well as $\gamma_j^c(m, m')$, the DTT.¹⁵

The key to identifying the TT effect of welfare (6) and assuring that the decomposition in (7) is correct, is the identification in the first step of the counterfactual expected outcome that individuals

that enter welfare would have had if they had not entered welfare. If the identity of the individuals on welfare that did not received any training could be observed, this would be unnecessary, because the end is to calculate the differential treatment effects of the training programs for welfare recipients only. However, it is necessary to take into account how the observed outcomes for welfare recipients compare to the outcomes of individuals not in welfare, to avoid confounding factors like the effects of the economic cycle or the changes in characteristics of new entrants. This makes indispensable to estimate the TT effect of welfare entry. The assumptions needed to identify this effect will be discussed below, but first it is necessary to take into account the consequences to the analysis above if the treatment effects of the training programs are not homogeneous.

3.3 Heterogeneous treatment effects

If the treatment on the treated effects are heterogeneous, then the identification of the TT and DTT becomes more difficult. It is still true that the TT effect of welfare can be decomposed as

$$\begin{aligned}\Delta_j^{ct} &= \sum_m P_m^{ct} E[Y_{it+j}^{ct}(m) - Y_{it+j}(0) | W_i^{ct} = 1, T_i^{ct}(m) = 1] \\ &= \sum_m P_m^{ct} \theta_j^{*ct}(m)\end{aligned}\tag{8}$$

where $\theta_j^{*ct}(m)$ is the Conditional Average Treatment on the Treated effect of treatment m , CTT, as defined in (2).

In a second step the variation across cohorts in the probabilities of training can be again exploited to identify the CTT effects. As in the previous section it is necessary to assume that the CTT effects are constant over time in a county,¹⁶ that is $\theta_j^{*ct}(m) = \theta_j^{*ct'}(m) = \theta_j^{*ct''}(m) = \theta_j^{*c}(m)$.

Identifying the CTT effects of each of the m treatments is interesting in itself, but does not allow to identify the DTT effects $\gamma(m, m')$, which are the central interest of this paper. However, it would still possible to identify these effects if there were variables available that explain the changes for different cohorts in the probabilities of receiving training, for reasons not related to the characteristics of the individuals in these cohorts. This can be seen clearly by decomposing the effect of treatment m on individual i in two parts,

$$\begin{aligned}
\theta_{ij}^{ct}(m) &= Y_{it+j}^{ct}(m) - Y_{it+j}(0) \\
&= \theta_j^{ct}(m) + \delta_{ij}^{ct}(m)
\end{aligned}$$

where $\theta_j^{ct}(m)$ represents the TT effect of treatment m , as defined in (1) and $\delta_{ij}^{ct} = \theta_{ij}^{ct} - \theta_j^{ct}$ represents the deviation respect to the Average TT for individual i . Then, the CTT effect of treatment m , identified in a second step as before, can be decomposed as

$$\theta_j^{*c}(m) = \theta_j^c(m) + E[\delta_{ij}^c(m)|W_i^{ct} = 1, T_i^{ct}(m) = 1]. \quad (9)$$

Replacing (9) in (8) and defining $\eta_j^c(m) = E[\delta_{ij}^c(m)|W_i^{ct} = 1, T_i^{ct}(m) = 1]$, the following expression is obtained:

$$\Delta_j^{ct} = \sum_m P_m^{ct}[\theta_j^c(m) + \eta_j^c(m)] \quad (10)$$

where the second term in the bracket represents the heterogeneity in treatment effects across different cohorts due to the average observed and unobserved characteristics of the individuals in each cohort. This can be interpreted as a problem of endogeneity where if appropriate instruments are found, then the TT effects of interest ($\theta_j^c(m)$) will be identified. Such instruments need to satisfy the condition that they affect the probability of each cohort of receiving each type of training, but are not related to the observed or unobserved average characteristics of the individuals in the cohort. The availability of instruments that satisfy these conditions will be discussed in Section 5.

3.4 Identification of TT effect of welfare entry

So far I have refrained from discussing the conditions under which the Average Treatment on the Treated effect of welfare entry can be identified. The non-experimental literature on program evaluation provides different alternatives, which depend on the assumptions made about the nature of the selection process into welfare and training and about the outcome generation process (see for example Heckman, Lalonde and Smith, 1998).

Unconfoundness The most widely used assumption is unconfoundness or versions of it. Under this assumption, conditional on observed characteristics X , selection into training and into the

different treatments is independent of the potential outcomes. That is,

$$Y_{it+j}(0) \perp W_i | X_i \quad (\text{A})$$

$$Y_{it+j}(0) \perp T_i(m) | X_i. \quad (\text{B})$$

¹⁷It is easy to show that in this case by just calculating the difference of conditional on X means of observed outcomes for individuals entrants on welfare at time t and a control group, the decomposition in equation (7) will hold. To see that, take equation (6) and rewrite it using (4), as

$$\begin{aligned} \Delta_j^{ct} &= E[Y_{it+j}^{ct}(W) | W_i^{ct} = 1, X] - E[Y_{it+j}(0) | W_i = 0, X] \\ &= E[\sum_m T_i^{ct}(m) Y_{it+j}^{ct}(m) | W_i^{ct} = 1, X] - E[Y_{it+j}(0) | W_i^{ct} = 1, X] \\ &= \{ \sum_m P_m^{ct} E[Y_{it+j}^{ct}(m) | W_i^{ct} = 1, T_i^{ct}(m) = 1, X] \} - E[Y_{it+j}(0) | W_i^{ct} = 1, X] \\ &= \sum_m P_m^{ct} E[Y_{it+j}^{ct}(m) - Y_{it+j}(0) | W_i^{ct} = 1, T_i^{ct}(m) = 1, X] \\ &= \sum_m P_m^{ct} \theta_j^{ct}(m) \end{aligned} \quad (11)$$

where the second line is implied by the unconfoundness assumption (A), the third line is obtained by using conditional expectations, and the fourth line is implied by assumption (B).

Allowing for unobservable factors If selection into welfare and into different treatments occurs not only on observable variables, but also on unobservable factors then the unconfoundness assumption will not identify adequately the counterfactuals of interest. One way of dealing with this problem is to use data prior to the treatment period to “difference out” any unobservable factor that remains constant over time. To allow for selection on (time invariant) unobservables and also for time variant potential outcomes under no treatment, it is possible to use a control group, and assume that the changes over time in the outcome for the control group reflect the changes that the treated individuals would have experienced if they had not entered welfare (and received the different treatments). That is, it can be assumed that

$$\begin{aligned}
E[Y_{it+j}^{ct}(0) - Y_{it-l}^{ct}(0)|W_i^{ct} = 1, T_i^{ct}(m) = 1] \\
&= E[Y_{it+j}^{ct}(0) - Y_{it-l}^{ct}(0)|W_i^{ct} = 1] \\
&= E[Y_{it+j}(0) - Y_{it-l}(0)|W_i = 0]. \quad (12)
\end{aligned}$$

Then, the TT effect of welfare entry can be obtained by using a difference in difference strategy where the differences over time for the control group are subtracted from the differences over time for the treatment group,

$$\Delta_j^{ct} = E[Y_{it+j}^{ct}(W) - Y_{it-l}^{ct}(0)|W_i^{ct} = 1] - E[Y_{it+j}(0) - Y_{it-l}(0)|W_i = 0].$$

Although this might seem different to the unconfoundness assumption, if the variable of interest is redefined as $Y_{it+j}^{ct}(W) - Y_{it-l}^{ct}(0)$, then essentially it is the same situation than under unconfoundness, where conditioning on a sufficiently rich set of covariates, it is assumed that the evolution over time of the counterfactual for the treatment group is the same as the evolution over time of the outcome for the control group.

The role of the propensity score Defining $e(X) \equiv \Pr(W = 1|X) = E[W|X]$ as the conditional probability or *propensity score* of welfare entry, Rosembaum and Rubin (1983) show that if the unconfoundness assumption holds conditional on X , then it is also true conditional on $e(X)$. This reduction in dimensionality will be used when comparing individuals treated (welfare entrants) versus potential controls, by looking at their propensity score instead of the whole set of covariates X .

4 Data

The empirical analysis is based on county level information, and on individual level data on welfare entrants. The data on training is constructed as quarterly averages from published county level monthly reports by every county of California on the number of people participating in the GAIN and WTW programs, and in each of the activities of the program.¹⁸ Based on these reports,

the probabilities of training and participating in activities were constructed.¹⁹ The period under analysis is from 1995 Q1 to 1999 Q2. I use data only for the 25 biggest counties in California, to assure that there is a minimum number of entrants per quarter (at least 100 entrants per quarter).

The individual level data comes from two administrative datasets for the State of California (MEDS and UI base wage files) which provide some demographic and family information and detailed monthly welfare use histories on every individual ever in the welfare system in California, as well as quarterly earnings histories (before, during, and after welfare) for these individuals, as long as their jobs are covered under the Unemployment Insurance system (around 90% of the employment of the State).²⁰ Using the MEDS dataset, new entrants to welfare (defined as individuals who enter welfare for the first time as adults since January 1987) were identified.

Some individuals had to be dropped from the analysis sample because they had missing demographic information, or they belonged to a case which had characteristics that did not allow to construct reliable family structure variables.²¹ An additional sample restriction is that only adults between 18 and 45 years old are analyzed (with older welfare recipients it is not clear whether counties assign them to training or not, and if they do, how effective can training be).²²

Table 1 presents descriptive statistics for different entry cohorts to welfare, defined by year of entry (however, the empirical analyses will be based on quarter of entry). As it can be noted, the number of new entrants decreases over time, which is consistent with the significant reduction of the welfare caseload in the period. The number of entrants in 1995 is the highest one (over 100,000 individuals), but in the analyses complete employment and earning histories are available only for a (randomly selected) sample. All the analyses will account properly for the sampling scheme by which that random sample was obtained. The Table shows some changes in the demographic composition of the entrants (small increase in the proportion of blacks and decrease in the proportion of Hispanics, and Spanish speakers), and in the composition of the families entering (the number of children at entry decreased over time, and the proportion of entrants with infants increased). Also, there are some differences in the employment and earning histories before entry, with a downward tendency in both employment and earnings measured in quarters 8 and 12 before entry.

Other variables that will be used in the regression analyses are measures of economic conditions at the county level. They include unemployment rate, employment to population ratio (total and in different sectors), measures of average earnings for all the workers and for specific sectors, as well as growth rates of average employment and earnings.²³

5 Estimation strategy

The estimation is based in the two-step strategy presented in section 3.

First step The first step entails estimating the treatment effect of welfare entry. The key for this is the availability of a control group not subject to treatment that satisfies the identifying assumptions. The data available is comprised only of welfare recipients, but they enter for the first time to welfare at different points in time. Then, these future cohorts of entrants, can be used as control groups for periods prior to entering welfare. The biggest problem in using future cohorts of entrants, is that it is necessary to be careful with how many periods before welfare entry are admissible to be used, because of the possibility that employment and earnings decline during some periods prior to the program entry. This phenomenon known in the literature as the “Ashenfelter Dip” is present in the data under analysis. This is solved by imposing a limit on how many periods before entering into welfare a future entrant can be used as a control (up to five quarters before entry).

The most relevant issue in the non-experimental evaluation literature is how to select the control group that satisfies the identifying assumptions. I use the propensity score (probability of welfare entry) as the basis of comparison between individuals. To estimate the propensity score, all time periods are re-expressed respect to the welfare entry quarter (time 0). An individual can be used as control for several different treatment cohorts, as long as he or she satisfies the specified condition of being at least five quarters apart from his or her own entry to welfare.²⁴ Because time 0 is redefined for each entry cohort, even when an individual is used more than one time, his or her information used varies according to the entry cohort to which is being compared. Pooling together

all the observations generated in this way (where all the observations are expressed in terms of time 0), the propensity score can be estimated for each observation. The estimation of the propensity score is done by running a logit model for welfare entry, in which the covariates are individual characteristics (the ones presented in Table 1), measures of local economic conditions (for the calendar quarter corresponding to time 0), full earnings and employment histories in the twelve quarters preceding to time 0, and indicator functions for county and cohort.

The advantage of estimating a propensity score model with the pooled observations is that a common support condition can be imposed across entry cohorts. In particular, I am interested not only in comparing treated individuals with control individuals of similar characteristics (as measured by the propensity score), but also it is important to impose that the treatment cohorts be comparable over time. Satisfying this second condition increases the credibility of the second step in the estimation of the differential treatment effects of training.

After imposing the common support condition across treatment cohorts,²⁵ the propensity score is re-estimated but separately for each entry cohort and its potential controls, and a common support condition is imposed between the treated and control individuals in each cohort.²⁶ The propensity score is re-estimated within entry cohorts for the remaining individuals.

Once the common support conditions are imposed, four set of estimators can be calculated for each outcome variable in each quarter after welfare entry. The first two are OLS estimators of the treatment effect of welfare entry, not-adjusting for covariates, and adjusting for covariates. The covariates-adjusted equation estimated is (the unadjusted one assumes $\beta_{2t}^c = \beta_{3t}^c = 0$)

$$Y_{it}^c = \beta_{0t}^c + \beta_{1t}^c W_i^c + \beta_{2t}^c X_i + \beta_{3t}^c W_i^c (X_i - \bar{X}_W) + \varepsilon_{it} \quad (13)$$

where Y_{it}^c represents the outcome of individual i in county c at period t , W_i^c is an indicator if individual i entered welfare at period 0 (before t), X_i is a set of covariates (which can include lagged values of the outcome variable), and \bar{X}_W represents the mean of the X_i variables for treated individuals ($\bar{X}_W = \Sigma W_i^c X_i / \Sigma W_i^c$). Then, β_{1t}^c represents the TT effect in period t of entering welfare in period 0 (before t), for individuals that entered welfare at county c . That is, $\beta_{1t}^c = \Delta_j^{ct}$

in terms of the notation in section 3.

The second set of estimators use the propensity score in estimating the treatment effects, as a way to select the appropriate controls for the treated individuals.²⁷ Following Hirano, Imbens and Ridder (2000) and Hirano and Imbens (2002), the propensity score will be used to re-weight the individuals in the control group. For estimating TT effects this implies constructing weights for each individual, based on the estimated propensity score, of the form: $\omega_i = W_i + (1 - W_i) [\hat{e}(Z)/1 - \hat{e}(Z)]$, where again W_i represents an indicator for treated individuals and $\hat{e}(Z)$ represents the estimated propensity score for individual i based on covariates Z . Using these weights an equation equivalent to (13), with or without extra covariates, can be estimated by Weighted Least Squares, using ω_i as the weight for each observation.²⁸

In all these cases a diff-in-diff estimator will be implemented by defining a new variable as the difference of the outcome of interest respect to a lagged value of the outcome before welfare entry.

Second step In the second step, $\hat{\beta}_{1t}^c$, the coefficient estimated in equations of the form (13) in the first step, can be used as a dependent variable in a regression of the form

$$\hat{\beta}_{1t}^c = \theta_t(L)P_{Lt}^{c0} + \theta_t(H)P_{Ht}^{c0} + \theta_t(N)P_{Nt}^{c0} + v_t^c$$

where $\theta_t(L)$, $\theta_t(H)$, and $\theta_t(N)$ represent the average conditional TT effects of LFA training, HCD training and no training, respectively, t periods after entering welfare, and P_{Lt}^{c0} , P_{Ht}^{c0} , and P_{Nt}^{c0} represent the probability at time 0 for individuals entering welfare in period t in county c of receiving LFA, HCD and no training respectively. As the sum of the probabilities of each treatment equals one, then $P_{Nt}^{c0} = 1 - P_{Lt}^{c0} - P_{Ht}^{c0}$, and it is possible to rewrite this equation as:

$$\hat{\beta}_{1t}^c = \alpha_{0t} + \alpha_{1t}P_{Lt}^{c0} + \alpha_{2t}P_{Ht}^{c0} + v_t^c \tag{14}$$

where α_0 represents the TT effect of no training, while α_1 and α_2 represent the TT effect of LFA and HCD training relative to no training.

There is a problem associated to the estimation of (14). As it was discussed in section 3, under heterogeneous treatment effects, α_0 , α_1 and α_2 will not represent the treatment effects of

interest, but conditional TT effects, because they will include the impact of the average observed and unobserved characteristics of the individuals in each cohort, on the average TT effects. In addition, not only the average characteristics of the cohort will determine the CTT, but also, as is shown in Mitnik (2005), the training policies will be endogenous to such characteristics, because the counties will adjust their policies to the individuals they need to serve. These two issues can be solved by using instrumental variables for the probability of LFA and HCD training. Mitnik (2005) shows that there are several variables that determine the training policies, and that are not related to the characteristics of the cohort treated. Those variables include total budget for the county, local economic conditions measures, and variables that are proxy for political participation in the county, like voting and registration patterns. These variables will be then used as instruments for P_{Lt}^{c0} and P_{Ht}^{c0} in (14).

6 Testing identifying assumptions

Pre-treatment tests The identifying assumptions of section 3 can not be tested directly, but they can be tested in indirect ways using pre-treatment data for the individuals under analysis. These tests were proposed by Heckman and Hotz (1989), and are based on the assumption that if the identifying assumptions hold for a group of individuals before treatment, then they should hold after treatment. In this way under unconfoundness the following condition should hold:

$$E[Y_{it-l}^{ct}(0)|W_i^{ct} = 1, X_i] = E[Y_{it-l}^{ct}(0)|W_i^{ct} = 0, X_i],$$

where $l \geq 2$, and a test of the null hypothesis that the treatment effect calculated on period $t - l + 1$ is zero should not be rejected.

For the difference in difference estimator, the hypothesis to test is that the treatment effect calculated as

$$E[Y_{it-l}^{ct}(W) - Y_{it-l-j}^{ct}(0)|W_i^{ct} = 1] - E[Y_{it-l}(0) - Y_{it-l-j}(0)|W_i = 0],$$

or in its conditional version, are zero, for $l \geq 1$ and $j > 0$.

Balancing tests Dehejia and Wahba (2002) and Smith and Todd (2003, forthcoming) remark the usefulness of “balancing tests” that can help in choosing the specification of the propensity score. The idea of these tests is that after conditioning on the propensity score, there should not be dependence of W on Z (the variables included in the estimation of the propensity score). That is,

$$E[W|Z, \Pr(W = 1|Z)] = E[W| \Pr(W = 1|Z)].$$

Here I will resort to a simple test of difference of means for treated and control individuals, weighting control individuals by the propensity score, with weights similar to the ones presented in the previous section. These differences of means should be statistically not significant.

7 Results

Propensity score/common support condition Before the first step (estimation of TT effects of welfare), it was necessary to select the individuals that would be controls, and impose the common support conditions described in section 5. Based on the pooled dataset of all treated and potential control individuals a “pooled” propensity score was calculated. Histograms of the estimated propensity scores, for treated and control individuals, are presented in Figure 2. As it can be seen in the Figure, the distribution for the controls is more skewed to the right than for the treated, implying that there are very few or no controls comparable to the treated individuals with higher propensity score values, and very few treated individuals for the lower values. The Figure shows also with the dashed lines the lower and higher bound of the propensity score that satisfy the common support condition across treatment cohorts.²⁹ These bounds implied that 15% in the lower tail and 10% in the upper tail of the propensity score distribution for treated individuals were dropped for not satisfying the common support condition. Among the controls the percentages were much higher (as expected), with 52% of the potential controls dropped because of a low propensity score and 1% because of a high propensity score. Although the number of treated individuals dropped is significant, this will make cross-treated cohorts comparisons more credible.³⁰

Table 2 compares the means of several covariates and pre-treatment outcomes for the treated

and controls (weighted), as a sort of “balancing test”. Although in most of the cases the differences of means appear as statistically significant, the actual differences are extremely small in all the cases except for the lagged outcomes (earnings and employments) in the last quarter before welfare entry, and for the number of kids at entry. Different specifications of the propensity score that were tested did not produce different results.³¹

After imposing this “pooled” common support condition, an extra condition was imposed, this time to make sure that treatment and controls within each county/cohort combination satisfy a common support condition. The application of this extra condition implied that on average an extra 6% of the treated cases and 25% of the remaining potential controls were dropped. In the end, for all the county/cohort/outcome quarter cells, the *average* number of treated and controls *per cell* that satisfied the common support conditions were 434 and 3756 respectively.³²

First step: TT effects of welfare The first step estimation results were obtained by running different specifications based on equation (13). Several outcomes are of interest: employment (indicator for positive earnings in the quarter), quarterly earnings, and differentiated versions of these two variables respect to a base period. To avoid using a quarter that could be considered part of the “Ashenfelter dip”, the fifth quarter before entry to welfare was selected as base period. For each of these outcome variables the effects from 1 quarter up to 16 quarters after welfare entry were analyzed (data on employment and earnings is available until the last quarter of year 2000, so for the last entry cohort studied, 1999 Q2, only 6 quarters were analyzed).

Table 3 presents the estimated TT effects of welfare entry, for the average of all the counties and entry cohorts (the effects were estimated separately by cohort/county/quarter cell, the displayed effects are weighted averages). Four specifications are shown for employment and four for earnings. The columns labeled “unweighted by the propensity score” refer to the estimation of (13) by OLS, while the ones labeled “weighted by the propensity score” estimate the same equation but with WLS, with weights for the control group as specified above. Within each of these specifications, the ones labeled “level” refer to the estimation with the outcome variable in levels including covariates in the regressions. The ones labeled “diff-in-diff” refer to the estimation with the outcome variable

differentiated respect to quarter 5 before welfare entry, and do not include other covariates.

The results in Table 3 show that in most of the cases there are not huge differences in the patterns of the coefficients across estimators, but there are differences in precision, with weighted diff-in-diff estimators showing the smaller standard errors. For those estimators both for employment and earnings, the initial effect of welfare entry appears to be negative or non significant becoming statistically significant (positive) from quarter 6 onwards.

The results in Table 3 are not very useful in the sense that they do not provide any guidance on how to choose among the four alternative estimators presented. The specifications tests presented in the previous section can be specially helpful with respect to this issue. Table 4 presents summary results of applying these tests to the four estimators under analysis. For the level estimators the test is performed by computing the treatment effect for lagged outcomes at 1, 4, 8 and 11 quarters before welfare entry. For the diff-in-diff estimators, the lagged outcomes are differentiated with respect to the previous period lagged outcome (i.e., quarter 1 is differentiated by quarter 2, etc.). This is done for each cohort/county combination, for the treatment and control groups used in the different periods after entry. The reason why it is necessary to perform the test for each period after welfare is that the control groups can (potentially) change each period, when individuals stop being eligible as controls due to being too close to their own welfare entry. The Table shows the proportion of county/cohort “cells” for which the test is statistically significant at the 5% significance level. At this significance level it should be expected no more than 5% of the cases in which the null hypothesis is rejected, but as it is clear from the table the rejection rate is higher than that in all the cases. The unweighted estimators perform particularly poorly, although the unweighted diff-in-diff estimator performs slightly better. For the weighted by propensity score estimators the diff-in-diff perform in almost all cases better than the level estimators. Respect to the evolution over quarters after welfare, for the first 6/8 quarters the diff-in-diff estimator performs relatively well, and the quality of the matches seem to worsen for the later quarters after entry. This is probably not surprising given that for the later periods the number of controls available gets smaller by definition. Overall the results in Table 4 suggest that the weighting by propensity

score diff-in-diff estimator should be the preferred one. Note, that the results do suggest that the estimator is not able to account for all the pre-entry differences between treatment and control groups. So, the results should be analyzed keeping this caveat in mind.

The treatment effect of entering welfare is interesting in itself, in particular when being analyzed, for example, by entry cohorts. Figures 3 to 6 present the treatment effects for the weighted by propensity score estimators in levels and diff-in-diff, for employment and earnings. Only the first cohort of each year is presented to make the figures readable. A clear pattern emerge: the earliest cohort (19995 Q1) had a much harder time recovering from the shock(s) that generated their entry to welfare (8 quarters until the employment effects become non-negative, and 6 quarters for earnings). The later cohorts have had much shorter periods of negative outcomes, and had very similar patterns of positive effects on employment and earnings from quarter 6 onwards.

What is the interpretation of these results? The central argument of this paper is that the differences in the effects of welfare entry reflect, at least in part, the policy changes in the training policies (GAIN and WTW Programs) followed by the counties. The next step is precisely estimating the effect of the training policy changes on the TT effects of welfare entry, and obtain the differential effects of LFA versus HCD programs.

Second step: conditional TT effects of training policies In this second step, the TT effects estimated using the weighting by propensity score diff-in-diff estimator in the first step are used as dependent variables in regressions like equation (14). For each outcome (employment or earnings) and for each quarter after entering welfare, WLS regressions were estimated, with the weights given by the average in the period of the welfare caseload of each county, to take into account the differences in sizes between the counties

Tables 5 and 6 show the results for employment and earnings respectively. Each table shows the results of estimating just by OLS, and of estimating an IV model in which $\Pr(LFA)$ and $\Pr(HCD)$ are instrumented. The choice of instruments is determined by the results in Mitnik (2005) who showed that variables not related to the characteristics of the individuals have an important effect on the training policies.³³ The statistics for the first stage of the IV regressions included in the

Tables show that the instruments are highly correlated with the probabilities of training variables and that they are jointly significant.

As it was explained in Section 5, the coefficients in regressions of the form (14) can be directly interpreted as the differential treatment effects of LFA and HCD versus no training. The results from the IV regressions are not drastically different from the OLS ones, although they are higher for the LFA coefficients, and in some cases statistically more significant for the no training and HCD coefficients. Both for employment and earnings the pattern that emerges is one in which for the first 9/10 quarters the effects of LFA training are strongly positive (around 20% for employment, between \$800 and \$1000 for earnings) and are non significant or negative for the fourth year after welfare entry.

The results on HCD training are mostly non significant or negative, while for no training, the results are negative the first few quarters, non significant up to the 7th quarter, and from the 8th quarter after entry they appear as positive and significant both for employment and earnings (the effects on employment are of around 8% and on earnings around \$500/600).

Tables 7 and 8 present again the IV specifications but including a dummy indicating the cohorts that entered welfare after welfare reform was implemented (first quarter of 1998 in California). Interactions of this dummy with the probabilities of LFA and HCD training were also included, to analyze if there are different treatment effects in the period after welfare reform. Under this specification it is very remarkable how the pre-welfare effects of LFA training become really big and positive, and in the post-welfare reform period they are close to zero. This suggest that after welfare reform the training programs became less significant and that the entry cohorts show a very fast recovery towards positive outcomes just 2 or 3 quarters after entry. The fact that after welfare reform time limits (that reduce the incentives for individuals to stay on welfare) and strongly financial incentives for work were introduced, could explain the turnaround.

In summary, the results are relatively in line with the findings from previous evaluations of WTW programs, with LFA training showing early positive results that fade after around 2 years and HCD training not showing positive results in the four years after entry to welfare.³⁴ When

allowing for differential effects for the pre and post welfare reform periods, the striking fact is that the previous results are stronger for the pre-welfare reform period, but for the post welfare reform period the training programs do not seem to have any positive effect.

8 Conclusions

In this paper I developed an econometric strategy that allowed me to estimate the effects of two different training strategies for welfare recipients, under a situation of unknown treatment status, using data for Welfare to Work program in California. This is achieved by first estimating the treatment effect of welfare entry, for different entry cohorts, using non-experimental methods based on propensity score weighting. And second, by exploiting cross county and cross time heterogeneity in training policies followed by the counties of California.

The results show that the effects of welfare entry change over time with later cohorts experiencing a much rapid recovery from the shock(s) that determine their entry in the first place. These changes are, at least partly, attributed to changes in training policies by counties in California.

By explaining the patterns of the effects of entry to welfare using the training policies, it is possible to obtain measures of the differential treatment effects of the LFA and HCD training strategies. The results are consistent with previous evaluations of these types of training programs, with LFA training showing positive effects on employment and earnings for around 10 quarters, and fading afterwards, and HCD training showing non-significant or negative effects. For the no training alternative the effects are interesting, because the effects become positive around 6 quarters after welfare entry.

A particularly surprising result is that if the treatment effects are allowed to differ before and after welfare reform, the pre-welfare reform results are similar to the ones before, but the after welfare reform results show no effect of none of the training strategies, and strong positive effect of the no training option. The implementation after welfare reform of time limits and work and financial incentives rules, among other changes, might have generated such changes in the entry and welfare use decisions of individuals that is very difficult to interpret this pattern. Also the

relatively short period of time after welfare reform covered by this study suggest that it would be necessary to extend the analysis further to obtain more clear patterns.

However, this result highlights an important limitation of the prior evaluations of training programs through experiments. Because none of the experimental studies were conducted in an environment like the one generated after welfare reform, it is very limited what can be learned from them in terms of applying similar programs to very different contexts.

If in addition one takes into account the cost and practical difficulties (for example how to select a randomized out control group in a world with time limits) of conducting experiments, the fact that a non-experimental methodology like the one proposed in this paper can be used to estimate the treatment effects of interest, is very encouraging.

Notes

¹The Job Opportunity and Basic Skills (JOBS) Program was directed to help families on welfare avoid long-term welfare use by providing job search assistance, education, work experience, vocational training, and other employment-related services, and required all parents (except those with small children) to participate in these work-related activities or face a reduction in the amount of assistance received (Haveman and Wolfe, 2000).

²The three sites were Riverside (California), Grand Rapids (Michigan) and Atlanta (Georgia).

³Grogger, Klerman and Karoly (2002, pp. 96) state: *“The results indicate that this type of policy generally increases both employment and earnings. In the short run, work programs that focus on job search generally yielded greater effects than programs that focus on skills building. After the first few years, the impacts of the program faded, although most remained positive. The impact differential between the search-oriented programs and the skills-oriented programs also faded with time”*.

⁴Grogger, Karoly and Klerman (2002) analyze a few experiments that mix work requirements and financial incentives or time limits, and although the results seem to be positive, they are not conclusive.

⁵The study by Hotz, Imbens and Klerman (2000), which is based on GAIN’s evaluation, raises another issue which is the long-term differential effects of LFA versus HCD programs. In their study they show that over a 9 year periods, the effects for Alameda and Los Angeles counties are similar to, or even slightly better than, the effects for Riverside.

⁶A different approach is proposed by Manski (2000, 2001) by using previous experimental results to inform the treatment decisions for different populations. Pepper (2002) applies this methodology using information for the WTW programs in the NEWWS experiment.

⁷Heckman and Smith (1995) provide several other disadvantages associated to conducting social experiments.

⁸TANF replaced AFDC after the passage of the Personal Responsibility and Work Opportunity Reconciliation Act (PRWORA) in 1996. California was the last state to implement TANF, starting in January 1998 under the CalWORKs name.

⁹However, participation in the WTW program does not mean necessarily receiving training services. There are non-training activities that count as participation in the WTW program (like using substance abuse services).

¹⁰This paper covers the period 1995-June 1999 because after June 1999 a new report system for the WTW program (WTW25) was implemented by the state of California, that was not fully functional until 2000. Because the comparability of the data from WTW25 with the data from the original report system (GAIN25) is far from clear, only the data from GAIN25 is used.

¹¹Figure 1 does not show the important cross county variation in these proportions, which will be exploited in the empirical application.

¹²The “no training” treatment might mean, then, that if the individual stayed long enough in welfare, she could have received eventually some training services, even though that did not happen as a new entrant.

¹³See for example Heckman and Hotz (1989), Friedlander and Robins (1995), Heckman and Smith (1995), Dehejia and Wahba (1998, 2002), Bloom et. al (2002), and Smith and Todd (2003).

¹⁴See Heckman, Lalonde and Smith (1998) on the uses and interpretation of the different parameters normally estimated in the evaluation literature.

¹⁵Alternatively, it could be assumed that the TTs are constant across counties, i.e. $\theta_j^{ct}(m) = \theta_j^{c't}(m) = \theta_j^{c''t}(m) = \theta_j^t(m)$ for at least three counties c , c' , and c'' for the cohort of entrants at time t , and $\theta_j^t(m)$ and $\gamma_j^t(m, m')$ will be identified.

¹⁶Or alternatively that they are constant across counties at a point of time.

¹⁷For the purposes of this paper the weaker assumption of conditional mean independence is sufficient. This implies that conditional on observable characteristics X , the expected potential outcomes are the same for every individual: $E[Y_{it+j}^{ct}(0)|W_i = 1, X] = E[Y_{it+j}^{ct}(0)|W_i = 0, X]$ and

$$E[Y_{it+j}^{ct}(m)|W_i = 1, T_i^{ct}(m) = 1, X] = E[Y_{it+j}(0)|W_i = 0, X_i].$$

¹⁸As has been already mentioned in section 2, the data used is from the GAIN25 report (that covers the GAIN Program period and the first year and half of the WTW Program).

¹⁹The monthly reports record the number of individuals participating in either the GAIN or WTW programs, and the number of “activities” offered during the month. Potentially some individuals might participate in more than one activity, which might imply that more activities are offered than people in the program. Then the proportions of individuals in LFA and HCD programs were calculated by summing up the activities in each category dividing by the sum of LFA and HCD total number of activities, and then multiplying this proportion by the ratio of the number of adults participating in the GAIN/WTW program to the total adult caseload, in the month. In some few cases (for small counties) this last ratio resulted in numbers slightly over one. In these cases the ratios were rounded to one.

²⁰The MEDS (MediCal Eligibility Data System) dataset goes from 1987 to 2000 and contains the welfare use and individual level information (although MEDS identifies welfare cases, no families, it is possible to use the case-level data to construct proxies for measures of family characteristics). The Unemployment Insurance (UI) base wage data goes from 1993 to 2000, and comes from the quarterly reports filed by employers to the California Employment Development Department (EDD). These two datasets were matched through the individuals Social Security Number, to form a dataset that contains not only the (monthly) welfare use history but also the (quarterly) earnings histories of every individual in Welfare in California during the 1990s.

²¹Missing or invalid personal information refers primarily to SSN not valid, and/or missing date of birth, sex or ethnicity. Also, because MEDS identifies cases, not families, it is necessary to be careful to identify cases that can be credibly called “families”. The cases dropped were all cases in which more than one adult in FG or two adults in UP cases received welfare, had more than two adults not receiving welfare, and had more than 5 kids on welfare.

²²In the 1995 Q1 through 1999 Q2 period there were 505,280 welfare entrants in California.

Of those, 147,331 (29.1%) were dropped from the analyses because they had missing or invalid information or they were older than 45 years. Of the remaining 357,949 cases, 30,601 (8.5%) were dropped because they entered welfare in small counties where less than 100 people entered welfare per quarter. The analysis sample is then composed of 327,348 new entrants for which their characteristics are presented in Table 2. In summary, 65% of all the new entrants in the period are included, with a concentration of half of these entrants in some few counties (Los Angeles, Orange, Riverside, San Bernardino, San Diego, and Sacramento). In addition to these, the following counties were included: Alameda, Butte, Contra Costa, Fresno, Imperial, Kern, Madera, Merced, Monterey, San Francisco, San Joaquin, Santa Barbara, Santa Clara, Shasta, Solano, Sonoma, Stanislaus, Tulare and Ventura

²³The Bureau of Labor and Statistics (BLS) publishes the county level unemployment rates, based on survey methods. The employment and earnings measures are published by the California Employment and Development Department (EDD) and BLS as part of the program known as ES-202 or Covered Employment and Wages.

²⁴When calculating the treatment effects for a given entry cohort, at different quarters after entry, the same individual may be used several times as a control as long as he or she are five quarters apart from their own entry.

²⁵The common support condition across entry cohorts requires that all individuals have a propensity score not lower than the highest of the minimum propensity score within each entry cohort, and not higher than the lowest maximum propensity score within each entry cohort.

²⁶The common support condition within an entry cohort requires that no control individual can have a lower propensity score than the lowest propensity score among treated individuals, and that no treated individual can have a higher propensity score than the highest propensity score among control individuals. In addition all individuals with propensity score below 0.02 and above 0.98 are dropped, which assures that no individual observation can have an excessive weight when weighting by the propensity score.

²⁷An alternative to use the propensity score would be to match individuals directly on the Xs , as proposed for example by Abadie and Imbens (2002).

²⁸An advantage of this method, as opposed to other methods like matching on the propensity score, for example the one proposed by Heckman, Ichimura and Todd (1998), is that it is less demanding computationally, which is important in this case given the large number cohorts for which TT effects need to be estimated.

²⁹This means that the $\max\{\text{minimum prop score}\}=0.1203$ and that the $\min\{\text{maximum prop score}\}=0.5545$.

³⁰Crump, Hotz, Imbens and Mitnik (2005) show how the application of an optimal common support condition changes the estimand (with respect to the original) but that this change will be desirable due to the efficiency gains associated to dropping individuals in the tails of the propensity score distribution.

³¹Also, it is interesting to compare the last column of Table 1 with the first column of Table 2, to assess the effects of applying the common support conditions, on the characteristics of the treated individuals.

³²Even after dropping a large number of controls due to the common support conditions, still a huge number of controls for each cohort/county/quarter remains available. This is particularly true for the earliest entry cohorts. Therefore, for the cases in which the remaining pool of potential controls is large, a random sample, by cohort/county/quarter “cell”, of those controls was selected.

³³The instruments used are $\log(\text{total budget})$, local economic conditions variables (employment/population ratio, average earnings, employment growth and average earnings growth for retail trade sector), political variables (voting and registration patterns for state assembly) and population-level demographic variables of the counties. See Mitnik (2005) for details.

³⁴Hotz, Imbens and Klerman (2005) find that the effects of HCD take a relative long time to show in the data and that over an extended period they can surpass the effects of LFA programs. So, is not surprising that in the two years after entry the effects of HCD are not significant.

References

- [1] Abadie, Alberto and Guido Imbens (2002). “Simple and Bias-Corrected Matching Estimators for Average Treatment Effects”, (August). Mimeo.
- [2] Blank, Rebecca, David Card, and Philip Robins (1999). “Financial Incentives for Increasing Work and Income Among Low-Income Families”. Working Paper #6998, National Bureau of Economic Research, (March).
- [3] Bloom, Howard, Charles Michalopoulos, Carolyn J. Hill, and Ying Lei (2002). “Can Non-experimental Comparison Group Methods Match the Findings from a Random Assignment Evaluation of Mandatory Welfare-to-Work Programs?”, Manpower Demonstration Research Corporation Working Paper (June).
- [4] Crump, Richard, V. Joseph Hotz, Guido Imbens and Oscar Mitnik (2005). ““Moving the Goalposts: Addressing the Limited Overlap in Estimation of Average Treatment Effects by Changing the Estimand” (November), mimeo.
- [5] Dehejia, Rajeev and Sadek Wahba (1998). “Causal Effects in Non-Experimental Studies: Re-evaluating the Evaluation of Training Programs”. Working Paper #6586, National Bureau of Economic Research, (June).
- [6] Dehejia, Rajeev and Sadek Wahba (2002). “Propensity Score-Matching Methods For Nonexperimental Causal Studies”, Review of Economics and Statistics, February 2002, Volume 84, No1, pp. 151-161.
- [7] Dehejia, Rajeev (2003). “Was There a Riverside Miracle? A Hierarchical Framework for Evaluating Programs With Grouped Data”, Journal of Business & Economic Statistics, Volume 21, No. 1 (January), pp. 1-11.
- [8] Dehejia, Rajeev (forthcoming). “Practical Propensity Score Matching: A Reply to Smith and Todd”, Journal of Econometrics, forthcoming.

- [9] Freedman, Stephen, Daniel Friedlander, Gayle Hamilton, Jo Ann Rock, Marisa Mitchell, Jodi Nudelman, Amanda Schweder, and Laura Storto (2000). “Evaluating Alternative Welfare-to-Work Approaches: Two-Year Impacts for Eleven Programs”. Manpower Demonstration Research Corporation (June).
- [10] Friedlander, Daniel and Philip K. Robin (1995). “Evaluating Program Evaluations: New Evidence on Commonly Used Nonexperimental Methods”, *American Economic Review*, Volume 85, Issue 4 (September), pp. 923-937.
- [11] Friedlander, Daniel, David Greenberg, and Philip Robins (1997). “Evaluating Government Training Programs for the Economically Disadvantaged”, *Journal of Economic Literature*, Volume 35, (December), pp. 1809-1855.
- [12] Grogger, Jeffrey, Lynn A. Karoly, Jacob A. Klerman (2002). “Consequences of Welfare Reform: A Research Synthesis”. RAND DRU-2676-DHHS (July).
- [13] Hamilton, Gayle, Thomas Brock, Mary Farrell, Daniel Friedlander, and Kristen Harknett (1997). “Evaluating Two Welfare-to-Work Program Approaches: Two-Year Findings on the Labor Force Attachment and Human Capital Development Programs in Three Sites”. Manpower Demonstration Research Corporation (December).
- [14] Hahn, Jinyong (1998). “On the Role of the Propensity Score in Efficient Semiparametric Estimation of Average Treatment Effects”, *Econometrica*, Vol. 66, No. 2, 315-333.
- [15] Hamilton, Gayle, Stephen Freedman, et. al. (2001). “How Effective are Different Welfare-to-Work Approaches? Five-Year Adult and Child Impacts for Eleven Programs”. Manpower Demonstration Research Corporation (December).
- [16] Haveman, Robert and Barbara Wolfe (2000). “Welfare to Work in the U.S.: A Model for Other Developed Nations?”, *International Tax and Public Finance*, Volume 7, pp. 95-114.
- [17] Heckman J. and R. Robb (1985). “Alternative Methods for Evaluating the Impact of Interventions” in “Longitudinal Analysis of Labor Market Data”, J. Heckman and B. Singer (eds).

- [18] Heckman, J., and V. J. Hotz, (1989). “Alternative Methods for Evaluating the Impact of Training Programs”, *Journal of the American Statistical Association*, 84(408):862-880.
- [19] Heckman, J., H. Ichimura, and P. Todd (1998). “Matching as an Econometric Evaluation Estimator”, *Review of Economic Studies*, 65, 261-294.
- [20] Heckman, James, and Jeffrey Smith (1995). “Assessing the Case for Social Experiments”, *Journal of Economic Perspectives*, 9(2), pp. 85-110.
- [21] Hirano, Keisuke, Guido W. Imbens, and Gert Ridder (2000). “Efficient Estimation of Average Treatment Effects Using the Estimated Propensity Score”, NBER Technical Working Paper 251.
- [22] Hirano, Keisuke and Guido W. Imbens (2002). “Estimation of Causal Effects using Propensity Score Weighting: An Application to Data on Right Heart Catherization”. Forthcoming in *Health Services and Outcomes Research Methodology*.
- [23] Hotz, V. Joseph, Guido Imbens, and Julie Mortimer (1999). “Predicting the Efficacy of Future Training Programs Using Past Experiences”, Technical Working Paper #238, National Bureau of Economic Research (May).
- [24] Hotz, V. Joseph, Guido Imbens, and Jacob Klerman (2005). “Evaluating the Differential Effects of Alternative Welfare-to-Work Training Components: A Re-Analysis of the California GAIN Program”, (October), mimeo.
- [25] LaLonde, Robert (1986). “Evaluating the Econometric Evaluation of Training Programs with Experimental Data”, *American Economic Review*, Volume 76, Issue 4 (September), pp. 604-620.
- [26] Manski, Charles (2000). “Identification problems and decisions under ambiguity: Empirical analysis of treatment response and normative analysis of treatment choice”. *Journal of Econometrics* 95, 415-442.

- [27] Manski, Charles (2001). “Statistical Treatment Rules for Heterogeneous Populations”. Mimeo (October).
- [28] Manpower Demonstration Research Corporation (2000). “Four-Year Impacts of Ten Programs on Employment Stability and Earnings Growth”. National Evaluation of Welfare-to-Work Strategies, (December).
- [29] Michalopoulos, Charles and Christine Schwartz (2000). “What Works Best for Whom: Impacts of 20 Welfare-to-Work Programs by Subgroup”. Manpower Demonstration Research Corporation (August).
- [30] Mitnik, Oscar (2005). “How Do Training Programs Assign Participants to Training? Characterizing the Optimal Assignment Rules of Government Agencies for Welfare-to-Work Programs in California”, mimeo.
- [31] Pepper, John (2002). “To Train or Not To Train: Optimal Treatment Assignment Rules Using Welfare-to-Work Experiments”. Mimeo (February).
- [32] Rosembaum, Paul and Donald Rubin (1983). “The Central Role of the Propensity Score in Observational Studies for Causal Effects”, *Biometrika*, Volume 70, No. 1 (April), pp. 41-55.
- [33] Smith, Jeffrey and Petra Todd (2003). “Does Matching Overcome Lalonde’s Critique of Non-experimental Estimators?” (November). Mimeo.
- [34] Smith, Jeffrey and Petra Todd (forthcoming). “Rejoinder”, *Journal of Econometrics*, forthcoming. [Response to Dehejia’s response to Smith and Todd, 2003]

**Table 1. Characteristics of individuals analyzed by year of first entry to welfare as adult
(Average 25 counties, standard deviations in parentheses)**

Variable	1995	1996	1997	1998	1999*	Avg. 1995-99
Female	0.76 (0.43)	0.74 (0.44)	0.75 (0.43)	0.74 (0.44)	0.75 (0.43)	0.75 (0.43)
White	0.35 (0.48)	0.36 (0.48)	0.37 (0.48)	0.34 (0.48)	0.34 (0.47)	0.35 (0.48)
Hispanic	0.40 (0.49)	0.39 (0.49)	0.38 (0.48)	0.37 (0.48)	0.37 (0.48)	0.39 (0.49)
Black	0.17 (0.37)	0.16 (0.37)	0.19 (0.39)	0.20 (0.40)	0.20 (0.40)	0.18 (0.38)
English	0.82 (0.38)	0.82 (0.39)	0.84 (0.37)	0.85 (0.36)	0.84 (0.36)	0.83 (0.38)
Spanish	0.12 (0.33)	0.12 (0.33)	0.10 (0.29)	0.08 (0.28)	0.09 (0.28)	0.11 (0.31)
Age at entry	26.0 (0.62)	26.7 (0.71)	26.2 (0.67)	25.8 (0.63)	25.6 (0.58)	26.2 (0.66)
FG/0P/1P at entry	0.72 (0.47)	0.69 (0.44)	0.71 (0.46)	0.69 (0.48)	0.71 (0.48)	0.71 (0.47)
Received welfare as child	0.34 (0.07)	0.25 (0.06)	0.30 (0.01)	0.35 (0.01)	0.37 (1.00)	0.32 (0.04)
Number of kids at entry	1.38 (0.43)	1.38 (0.43)	1.36 (0.44)	1.34 (0.45)	1.29 (0.45)	1.36 (0.44)
Kids<1 at entry	0.25 (0.50)	0.25 (0.50)	0.27 (0.50)	0.29 (0.50)	0.29 (0.50)	0.26 (0.50)
Kids 1-5 at entry	0.50 (0.43)	0.49 (0.44)	0.50 (0.43)	0.49 (0.41)	0.48 (0.41)	0.49 (0.43)
Kids 6-12 at entry	0.24 (0.30)	0.26 (0.32)	0.24 (0.30)	0.22 (0.29)	0.21 (0.27)	0.24 (0.30)
Kids 13-17 at entry	0.10 (0.45)	0.11 (0.46)	0.10 (0.46)	0.09 (0.46)	0.08 (0.45)	0.10 (0.46)
Employed 1 qtr before entry	0.32 (0.47)	0.34 (0.47)	0.33 (0.47)	0.33 (0.47)	0.34 (0.47)	0.33 (0.47)
Employed 4 qtrs before entry	0.35 (0.47)	0.37 (0.48)	0.36 (0.48)	0.35 (0.48)	0.36 (0.49)	0.36 (0.48)
Employed 8 qtrs before entry	0.33 (0.48)	0.35 (0.48)	0.33 (0.48)	0.32 (0.48)	0.31 (0.49)	0.33 (0.48)
Employed 12 qtrs before entry	0.29 (0.48)	0.31 (0.48)	0.29 (0.48)	0.27 (0.48)	0.26 (0.48)	0.29 (0.48)
Earnings 1 qtr before entry (\$1,000)	0.74 (0.48)	0.80 (0.48)	0.75 (0.48)	0.74 (0.48)	0.76 (0.48)	0.76 (0.48)
Earnings 4 qtrs before entry (\$1,000)	1.03 (0.47)	1.07 (0.48)	0.99 (0.48)	0.94 (0.47)	0.95 (0.48)	1.01 (0.48)
Earnings 8 qtrs before entry (\$1,000)	1.09 (0.47)	1.14 (0.48)	1.01 (0.48)	0.93 (0.47)	0.88 (0.47)	1.05 (0.47)
Earnings 12 qtrs before entry (\$1,000)	1.09 (0.47)	1.10 (0.48)	0.97 (0.47)	0.84 (0.47)	0.79 (0.46)	1.01 (0.47)
Number of Entrants**	35,525	84,188	63,492	54,434	24,822	262,461

Notes:

* First two quarters only

** For 1995 the actual number of entrants was 100,412 but only a random sample, for which lagged earnings and employment measures are available, is used. Thus the information for 1995 is weighted to account for the sampling scheme. For Avg. 1995-99 represents the total number of entrants analyzed in the period.

Table 2. Characteristics of pooled cohorts of treated vs. controls individuals that satisfy common support condition /a

Variable	Treated Mean	Controls Mean /b	Difference of Means
Female	0.77 (0.42)	0.76 (0.43)	0.01 ***
White	0.36 (0.48)	0.36 (0.48)	0.00 ***
Hispanic	0.38 (0.49)	0.38 (0.49)	0.00 **
Black	0.17 (0.38)	0.18 (0.38)	-0.01
English	0.85 (0.36)	0.84 (0.37)	0.01 ***
Spanish	0.10 (0.30)	0.11 (0.31)	-0.01 ***
Age at entry	26.0 (7.7)	26.5 (7.7)	-0.5 ***
FG/0P/1P at entry	0.72 (0.45)	0.72 (0.45)	0.00 ***
Received welfare as child	0.30 (0.46)	0.29 (0.45)	0.01 ***
Number of kids at entry	1.22 (0.99)	1.31 (1.05)	-0.09 ***
Kids<1 at entry	0.29 (0.45)	0.29 (0.45)	0.00 ***
Kids 1-5 at entry	0.43 (0.50)	0.48 (0.50)	-0.05 ***
Kids 6-12 at entry	0.22 (0.41)	0.25 (0.44)	-0.03 ***
Kids 13-17 at entry	0.11 (0.31)	0.11 (0.32)	0.00 ***
Employed 1 qtr before entry	0.32 (0.47)	0.33 (0.47)	-0.01 ***
Employed 4 qtrs before entry	0.37 (0.48)	0.36 (0.48)	0.01 ***
Employed 8 qtrs before entry	0.35 (0.48)	0.34 (0.47)	0.01 ***
Employed 12 qtrs before entry	0.30 (0.46)	0.30 (0.46)	0.00 ***
Earnings 1 qtr before entry (\$1,000)	0.63 (1.33)	0.74 (1.48)	-0.11 ***
Earnings 4 qtrs before entry (\$1,000)	0.98 (1.96)	1.00 (2.02)	-0.02 ***
Earnings 8 qtrs before entry (\$1,000)	1.04 (2.20)	1.04 (2.21)	0.00
Earnings 12 qtrs before entry (\$1,000)	1.01 (2.26)	1.01 (3.15)	0.00
Number of individuals	197,728	613,740	-

Notes:

*, **, *** denote difference statistically significant at 1%, 5% and 1% level respectively.

/a The propensity score was calculated pooling treated and control individuals for all the cohorts, and keeping only individuals with propensity score higher than the maximum of the minimum propensity scores across cohorts, and smaller than the minimum of the maximum propensity scores across cohorts (interval [0.1203,0.5545]).

/b The means for control individuals were calculated weighting by $(P_i/1-P_i)$, where P_i =propensity score for individual i .

Table 3. Average Treatment on the Treated Effects of first time welfare entry for 1995 Q1 to 1999 Q2 cohorts in 25 CA counties (weighted average of cohort/county treatment effects)

Qtrs After Welfare Entry	Number of Cohorts & Counties	Employment			Earnings (\$1,000)				
		Unweighted by Prop. Score		Weighted by Prop. Score		Unweighted by Prop. Score		Weighted by Prop. Score	
		Level	Diff-in-Diff	Level	Diff-in-Diff	Level	Diff-in-Diff	Level	Diff-in-Diff
1	450	-0.06 (0.02)	-0.10 (0.02)	-0.06 (0.01)	-0.07 (0.01)	-0.32 (0.08)	-0.44 (0.08)	-0.31 (0.04)	-0.29 (0.05)
2	450	-0.01 (0.06)	-0.07 (0.02)	-0.01 (0.02)	-0.03 (0.01)	-0.09 (0.21)	-0.26 (0.08)	-0.09 (0.06)	-0.12 (0.06)
3	450	0.01 (0.02)	-0.04 (0.02)	0.01 (0.01)	0.00 (0.02)	-0.01 (0.09)	-0.12 (0.06)	0.00 (0.05)	0.00 (0.06)
4	450	0.02 (0.04)	-0.01 (0.02)	0.02 (0.03)	0.02 (0.02)	0.15 (0.29)	0.06 (0.26)	0.17 (0.40)	0.15 (0.36)
5	450	0.04 (0.04)	0.01 (0.03)	0.04 (0.04)	0.03 (0.02)	0.17 (0.19)	0.11 (0.11)	0.18 (0.20)	0.19 (0.09)
6	450	0.05 (0.05)	0.02 (0.03)	0.05 (0.06)	0.05 (0.02)	0.26 (0.24)	0.21 (0.12)	0.27 (0.27)	0.27 (0.11)
7	450	0.06 (0.05)	0.03 (0.03)	0.06 (0.06)	0.06 (0.03)	0.31 (0.25)	0.27 (0.13)	0.27 (0.29)	0.33 (0.11)
8	450	0.06 (0.04)	0.03 (0.03)	0.06 (0.04)	0.06 (0.03)	0.33 (0.19)	0.31 (0.13)	0.31 (0.20)	0.34 (0.12)
9	425	0.07 (0.04)	0.03 (0.03)	0.07 (0.04)	0.07 (0.03)	0.38 (0.19)	0.36 (0.14)	0.40 (0.19)	0.39 (0.12)
10	400	0.07 (0.04)	0.04 (0.03)	0.08 (0.04)	0.07 (0.03)	0.41 (0.19)	0.40 (0.14)	0.42 (0.20)	0.44 (0.13)
11	375	-0.17 (2.37)	0.04 (0.03)	-0.17 (0.79)	0.08 (0.03)	-0.05 (11.86)	0.44 (0.14)	-0.03 (4.04)	0.50 (0.13)
12	350	0.08 (0.04)	0.04 (0.03)	0.07 (0.05)	0.08 (0.03)	0.48 (0.22)	0.49 (0.15)	0.47 (0.25)	0.52 (0.14)
13	325	0.06 (0.04)	0.03 (0.03)	0.06 (0.04)	0.07 (0.03)	0.44 (0.22)	0.46 (0.16)	0.41 (0.24)	0.51 (0.15)
14	300	0.08 (0.05)	0.04 (0.03)	0.08 (0.05)	0.08 (0.03)	0.51 (0.25)	0.48 (0.17)	0.47 (0.27)	0.53 (0.17)
15	275	0.07 (0.04)	0.04 (0.04)	0.07 (0.04)	0.08 (0.03)	0.53 (0.25)	0.54 (0.18)	0.50 (0.25)	0.54 (0.18)
16	250	0.06 (0.05)	0.04 (0.04)	0.04 (0.05)	0.07 (0.04)	0.50 (0.27)	0.58 (0.19)	0.45 (0.33)	0.54 (0.19)

Notes:

*, **, *** denote coefficient is significant at 10%, 5% and 1% level respectively (standard errors in parenthesis)
 The outcomes in "level" columns depict the results of regressions that include covariates in addition to the treatment indicator.
 The outcomes in "diff-in-diff" columns depict the results of regressions of the difference of the outcome variable respect to its value in quarter 5 before entry to welfare, without other covariates.
 The "unweighted by propensity score" specifications are regular OLS regressions, the "weighted by propensity score" ones are WLS regressions with weights for the control individuals given by $(P_i/1-P_i)$, where P_i is the propensity score for individual i .

Table 4. Proportion of pre-treatment outcomes statistically significant at 5% levels for 1995 Q1 to 1999 Q2 cohorts in 25 CA counties: (weighted average of cohort/county treatment effects)

Qtr After Entry in Which Treat/Cont Cohorts are Used	Number of Cohorts & Counties	Unweighted by Prop. Score					Weighted by Prop. Score										
		Level		Diff-in-Diff			Level		Diff-in-Diff								
		1	4	8	11	11	1	4	8	11	11						
Employment (Quarters Before Welfare Entry)																	
1	450	0.41	0.13	0.31	0.43	0.47	0.16	0.09	0.10	0.18	0.22	0.20	0.18	0.12	0.13	0.12	0.11
2	450	0.37	0.14	0.29	0.43	0.48	0.16	0.07	0.12	0.13	0.22	0.19	0.21	0.12	0.12	0.12	0.12
3	450	0.39	0.13	0.31	0.45	0.43	0.12	0.07	0.10	0.17	0.17	0.24	0.16	0.11	0.12	0.11	0.12
4	450	0.33	0.11	0.29	0.42	0.41	0.10	0.06	0.15	0.15	0.18	0.21	0.23	0.15	0.18	0.11	0.17
5	450	0.31	0.13	0.32	0.44	0.37	0.12	0.08	0.13	0.23	0.17	0.22	0.21	0.15	0.11	0.18	0.14
6	450	0.27	0.18	0.36	0.39	0.34	0.08	0.08	0.10	0.16	0.17	0.22	0.31	0.14	0.17	0.17	0.21
7	450	0.31	0.22	0.46	0.46	0.35	0.17	0.22	0.19	0.18	0.26	0.30	0.31	0.23	0.23	0.23	0.23
8	450	0.33	0.28	0.50	0.50	0.36	0.18	0.25	0.25	0.28	0.31	0.36	0.36	0.27	0.29	0.32	0.35
9	425	0.37	0.27	0.51	0.45	0.37	0.20	0.26	0.25	0.28	0.36	0.42	0.34	0.32	0.30	0.37	0.33
10	400	0.31	0.32	0.57	0.43	0.35	0.19	0.36	0.25	0.33	0.29	0.46	0.39	0.34	0.30	0.50	0.38
11	375	0.30	0.34	0.55	0.49	0.31	0.20	0.40	0.26	0.25	0.36	0.46	0.39	0.33	0.37	0.42	0.43
12	350	0.30	0.32	0.63	0.48	0.36	0.24	0.49	0.24	0.43	0.35	0.53	0.43	0.33	0.32	0.52	0.43
13	325	0.31	0.38	0.59	0.40	0.36	0.23	0.52	0.28	0.27	0.39	0.57	0.40	0.39	0.30	0.56	0.38
14	300	0.28	0.33	0.58	0.51	0.30	0.24	0.49	0.28	0.38	0.44	0.57	0.39	0.42	0.37	0.54	0.43
15	275	0.35	0.35	0.57	0.53	0.35	0.24	0.51	0.32	0.42	0.33	0.62	0.44	0.44	0.37	0.55	0.48
16	250	0.30	0.53	0.61	0.32	0.31	0.32	0.64	0.28	0.44	0.42	0.65	0.39	0.31	0.33	0.60	0.41
Earnings (Quarters Before Welfare Entry)																	
1	450	0.77	0.26	0.10	0.20	0.57	0.29	0.09	0.08	0.22	0.17	0.16	0.21	0.16	0.21	0.20	0.21
2	450	0.75	0.22	0.10	0.24	0.54	0.28	0.08	0.12	0.27	0.22	0.17	0.22	0.20	0.19	0.19	0.23
3	450	0.75	0.22	0.13	0.25	0.54	0.27	0.06	0.09	0.21	0.17	0.16	0.27	0.23	0.16	0.17	0.17
4	450	0.74	0.27	0.09	0.25	0.55	0.28	0.09	0.16	0.26	0.21	0.23	0.18	0.20	0.17	0.19	0.18
5	450	0.68	0.22	0.17	0.27	0.48	0.22	0.09	0.11	0.22	0.20	0.25	0.23	0.18	0.22	0.20	0.21
6	450	0.64	0.23	0.22	0.26	0.41	0.24	0.08	0.09	0.26	0.17	0.22	0.25	0.16	0.21	0.18	0.24
7	450	0.63	0.25	0.33	0.33	0.41	0.28	0.12	0.17	0.25	0.22	0.31	0.32	0.25	0.27	0.23	0.29
8	450	0.65	0.31	0.34	0.36	0.49	0.34	0.21	0.17	0.35	0.24	0.40	0.37	0.33	0.39	0.30	0.31
9	425	0.71	0.26	0.35	0.35	0.49	0.36	0.18	0.19	0.35	0.31	0.46	0.41	0.34	0.35	0.32	0.39
10	400	0.70	0.25	0.37	0.36	0.50	0.32	0.21	0.20	0.39	0.31	0.41	0.43	0.40	0.36	0.36	0.39
11	375	0.68	0.28	0.41	0.33	0.40	0.30	0.22	0.29	0.42	0.35	0.59	0.40	0.37	0.41	0.39	0.54
12	350	0.65	0.29	0.45	0.35	0.40	0.24	0.21	0.27	0.46	0.37	0.58	0.46	0.32	0.36	0.40	0.46
13	325	0.62	0.24	0.46	0.37	0.44	0.27	0.21	0.27	0.50	0.43	0.53	0.49	0.39	0.46	0.43	0.50
14	300	0.66	0.28	0.46	0.39	0.37	0.32	0.22	0.27	0.43	0.41	0.52	0.51	0.36	0.31	0.42	0.52
15	275	0.65	0.23	0.47	0.39	0.39	0.27	0.29	0.32	0.46	0.35	0.63	0.39	0.34	0.39	0.50	0.45
16	250	0.63	0.33	0.51	0.31	0.41	0.29	0.41	0.37	0.46	0.36	0.48	0.52	0.35	0.50	0.47	0.52

Notes:
 The "level" columns depict statistics for regressions that include covariates in addition to the treatment indicator.
 The "diff-in-diff" columns depict statistics for regressions of the difference of the variable respect to its value in the previous period.
 The "unweighted by propensity score" specifications are regular OLS regressions, the "weighted by propensity score" ones are WLS regressions with weights for the control individuals given by $(P_i/(1-P_i))$, where P_i is the propensity score for individual i .

Table 5. Differential Treatment Effects of training programs for 1995 Q1 to 1999 Q2 cohorts in 25 CA counties
Outcome: quarterly employment

Qtrs After Welfare Entry	Number of Observations	OLS			IV for Proportions LFA & HCD Training								
		No Training (NT)	LFA-NT	HCD-NT	P-Value F Test Coeffs=0	No Training (NT)	LFA-NT	HCD-NT	P-Value F Test Coeffs=0	R ² Instrum.	P-Val F Instr=0	R ² Instrum.	P-Val F Instr=0
1	441	-0.13*** (0.01)	0.08*** (0.01)	0.03 (0.03)	0.00	-0.13*** (0.01)	0.10*** (0.02)	0.00 (0.09)	0.00	0.83	0.61	0.83	0.00
2	441	-0.11*** (0.01)	0.12*** (0.02)	0.03 (0.04)	0.00	-0.11*** (0.01)	0.19*** (0.04)	-0.09 (0.11)	0.00	0.83	0.61	0.83	0.00
3	441	-0.07*** (0.02)	0.15*** (0.02)	0.02 (0.04)	0.00	-0.08*** (0.02)	0.23*** (0.04)	-0.14 (0.11)	0.00	0.83	0.61	0.83	0.00
4	441	-0.06*** (0.02)	0.18*** (0.03)	0.02 (0.05)	0.00	-0.07*** (0.02)	0.25*** (0.05)	-0.06 (0.13)	0.00	0.83	0.61	0.83	0.00
5	441	-0.02 (0.02)	0.18*** (0.03)	0.07 (0.05)	0.00	-0.04** (0.02)	0.26*** (0.04)	-0.02 (0.11)	0.00	0.83	0.61	0.83	0.00
6	441	0.00 (0.02)	0.17*** (0.03)	0.09* (0.05)	0.00	-0.01 (0.02)	0.23*** (0.05)	0.10 (0.12)	0.00	0.83	0.61	0.83	0.00
7	441	0.01 (0.02)	0.17*** (0.02)	-0.08* (0.04)	0.00	0.01 (0.02)	0.27*** (0.04)	-0.39*** (0.11)	0.00	0.83	0.61	0.83	0.00
8	441	0.05 (0.03)	0.04 (0.03)	-0.06 (0.08)	0.15	0.05* (0.03)	0.17*** (0.05)	-0.58*** (0.17)	0.00	0.83	0.61	0.83	0.00
9	416	0.05** (0.03)	0.03 (0.03)	-0.25*** (0.08)	0.01	0.06** (0.03)	0.18*** (0.05)	-0.85*** (0.17)	0.00	0.81	0.60	0.81	0.00
10	392	0.06*** (0.02)	0.01 (0.04)	-0.38*** (0.05)	0.00	0.06*** (0.02)	0.11* (0.06)	-0.84*** (0.19)	0.00	0.81	0.61	0.81	0.00
11	368	0.05* (0.03)	-0.03 (0.04)	-0.31*** (0.07)	0.00	0.07** (0.03)	0.02 (0.06)	-0.88*** (0.26)	0.01	0.82	0.63	0.82	0.00
12	344	0.05** (0.02)	-0.04 (0.06)	-0.25*** (0.08)	0.01	0.08*** (0.03)	-0.10 (0.09)	-0.88*** (0.29)	0.01	0.85	0.65	0.85	0.00
13	320	0.02 (0.03)	-0.22*** (0.08)	-0.25* (0.15)	0.04	0.06 (0.04)	-0.30** (0.12)	-1.10** (0.45)	0.02	0.88	0.76	0.88	0.00
14	296	0.06* (0.03)	-0.30*** (0.08)	-0.18 (0.12)	0.00	0.09** (0.04)	-0.32** (0.14)	-1.02** (0.44)	0.04	0.88	0.76	0.88	0.00
15	272	0.00 (0.03)	-0.32*** (0.09)	0.21 (0.27)	0.00	0.04 (0.03)	-0.53*** (0.16)	-0.49 (0.48)	0.01	0.89	0.88	0.89	0.00
16	248	0.01 (0.01)	-0.22* (0.11)	0.14 (0.24)	0.19	0.03 (0.03)	-0.43** (0.21)	-0.12 (0.58)	0.21	0.91	0.89	0.91	0.00

Notes:
*, **, *** denote coefficient significant at 10%, 5% and 1% level respectively (standard errors in parentheses, adjusted by county/fiscal year clusters).
All the regressions in this table include county fixed effects.
Each row represent a regression against LFA and HCD training proportions of the average treatment on the treated effects of welfare entry on employment. These treatment effects were estimated by taking differences of each employment outcome respect to its value on quarter 5 before entry, using WLS regressions with weights for the control individuals given by $(\pi_i/1-\pi_i)$, where π_i is the propensity score for individual i .

Table 6. Differential Treatment Effects of training programs for 1995 Q1 to 1999 Q2 cohorts in 25 CA counties
Outcome: quarterly earnings (\$1,000)

Qtrs After Welfare Entry	Number of Observations	OLS			IV for Proportions LFA & HCD Training								
		No Training (NT)	LFA-NT	HCD-NT	P-Value F Test Coeffs=0	No Training (NT)	LFA-NT	HCD-NT	P-Value F Test Coeffs=0				
1	441	-0.38*** (0.03)	0.03 (0.05)	0.07 (0.10)	0.00	-0.38*** (0.03)	-0.07 (0.10)	0.47 (0.37)	0.00	0.83	0.00	0.61	0.00
2	441	-0.26*** (0.06)	0.25*** (0.06)	0.24** (0.11)	0.00	-0.28*** (0.06)	0.35** (0.13)	0.21 (0.41)	0.00	0.83	0.00	0.61	0.00
3	441	-0.11* (0.06)	0.32*** (0.08)	0.28** (0.14)	0.00	-0.14** (0.06)	0.45*** (0.14)	0.25 (0.38)	0.00	0.83	0.00	0.61	0.00
4	441	0.04 (0.14)	0.02 (0.51)	0.91** (0.41)	0.00	-0.03 (0.13)	-0.37 (0.95)	3.16 (2.29)	0.01	0.83	0.00	0.61	0.00
5	441	0.25** (0.12)	0.58*** (0.12)	0.82*** (0.22)	0.00	0.20* (0.11)	0.84*** (0.19)	0.61 (0.49)	0.00	0.83	0.00	0.61	0.00
6	441	0.10 (0.07)	0.84*** (0.10)	0.62*** (0.13)	0.00	0.02 (0.08)	1.05*** (0.19)	0.86 (0.58)	0.00	0.83	0.00	0.61	0.00
7	441	0.19*** (0.04)	0.84*** (0.09)	0.44** (0.20)	0.00	0.17*** (0.05)	1.03*** (0.15)	0.06 (0.47)	0.00	0.83	0.00	0.61	0.00
8	441	0.34*** (0.10)	0.40*** (0.13)	-0.09 (0.34)	0.00	0.32*** (0.10)	1.18*** (0.32)	-2.56*** (0.88)	0.00	0.83	0.00	0.61	0.00
9	416	0.49*** (0.11)	0.13 (0.15)	-0.88** (0.37)	0.00	0.47*** (0.10)	1.00*** (0.28)	-3.84*** (0.97)	0.00	0.81	0.00	0.60	0.00
10	392	0.64*** (0.17)	0.03 (0.27)	-0.91 (0.56)	0.00	0.66*** (0.17)	0.48 (0.31)	-3.39*** (1.00)	0.00	0.81	0.00	0.61	0.00
11	368	0.54*** (0.11)	-0.28 (0.18)	-1.32** (0.58)	0.00	0.62*** (0.12)	-0.21 (0.25)	-3.29*** (1.02)	0.00	0.82	0.00	0.63	0.00
12	344	0.39*** (0.10)	-0.12 (0.26)	-1.36*** (0.50)	0.00	0.55*** (0.15)	-0.47 (0.44)	-4.09** (1.58)	0.00	0.85	0.00	0.65	0.00
13	320	0.30*** (0.11)	-0.83** (0.37)	-1.75* (0.99)	0.03	0.49*** (0.18)	-1.06 (0.71)	-6.60** (2.70)	0.04	0.88	0.00	0.76	0.00
14	296	0.63*** (0.17)	-1.22* (0.63)	-0.98 (0.85)	0.00	0.77*** (0.21)	-0.98 (1.06)	-5.34* (2.72)	0.00	0.88	0.00	0.76	0.00
15	272	0.32 (0.25)	-2.09*** (0.44)	0.93 (1.30)	0.00	0.44* (0.25)	-2.84*** (0.80)	-1.01 (2.10)	0.00	0.89	0.00	0.88	0.00
16	248	0.46*** (0.11)	-1.46* (0.81)	-0.06 (1.61)	0.00	0.59*** (0.22)	-2.55* (1.48)	-1.83 (3.74)	0.00	0.91	0.00	0.89	0.00

Notes:
*, **, *** denote coefficient significant at 10%, 5% and 1% level respectively (standard errors in parentheses, adjusted by county/fiscal year clusters).
All the regressions in this table include county fixed effects.
Each row represent a regression against LFA and HCD training proportions of the average treatment on the treated effects of welfare entry on earnings. These treatment effects were estimated by taking differences of each earnings outcome respect to its value on quarter 5 before entry, using WLS regressions with weights for the control individuals given by $(\pi_i/1-\pi_i)$, where π_i is the propensity score for individual i.

Table 7. Differential Treatment Effects of training programs for 1995 Q1 to 1999 Q2 cohorts in 25 CA counties including Welfare Reform indicator and interactions with training probabilities (IV specification)
Outcome: quarterly employment

Qtrs After Welfare Entry	Number of Observations	No Training (NT)	LFA-NT	HCD-NT	Welfare Reform Dummy	LFA - NT		HCD - NT		F Test Welfare Reform Interact. Coefs=0
						Welf Ref	X	Welf Ref	X	
1	441	-0.13*** (0.02)	0.30*** (0.07)	-0.46 (0.29)	0.03 (0.03)	-0.30*** (0.07)	0.63* (0.37)	0.00		
2	441	-0.11*** (0.02)	0.45*** (0.08)	-0.60 (0.37)	0.08** (0.04)	-0.43*** (0.09)	0.67 (0.44)	0.00		
3	441	-0.07*** (0.02)	0.49*** (0.08)	-0.84** (0.38)	0.07** (0.04)	-0.47*** (0.09)	0.95** (0.47)	0.00		
4	441	-0.06** (0.03)	0.66*** (0.12)	-1.02** (0.43)	0.09* (0.05)	-0.66*** (0.13)	1.28** (0.56)	0.00		
5	441	-0.04 (0.02)	0.72*** (0.13)	-0.86** (0.42)	0.08* (0.05)	-0.65*** (0.14)	1.13** (0.54)	0.00		
6	441	-0.07** (0.03)	0.83*** (0.15)	-0.04 (0.38)	0.13*** (0.05)	-0.65*** (0.14)	0.06 (0.52)	0.00		
7	441	-0.02 (0.03)	0.70*** (0.15)	-0.68* (0.38)	0.12** (0.05)	-0.55*** (0.15)	0.32 (0.53)	0.00		
8	441	0.02 (0.04)	0.62*** (0.13)	-0.85 (0.61)	0.13** (0.05)	-0.58*** (0.14)	0.27 (0.74)	0.00		
9	416	0.03 (0.03)	0.60*** (0.14)	-1.11** (0.48)	0.11* (0.06)	-0.58*** (0.16)	0.46 (0.66)	0.00		
10	392	0.07* (0.03)	0.58*** (0.16)	-1.49*** (0.55)	0.17*** (0.06)	-0.88*** (0.21)	1.10 (0.80)	0.00		
11	368	0.07* (0.04)	0.46** (0.20)	-1.49*** (0.48)	0.09 (0.06)	-0.81** (0.32)	1.45* (0.87)	0.09		
12	344	0.03 (0.04)	0.64*** (0.18)	-0.75 (0.67)	0.16 (0.10)	-1.02** (0.46)	0.14 (1.33)	0.00		
13	320	0.07 (0.05)	-0.06 (0.18)	-1.53** (0.71)	0.13 (0.16)	-1.28 (0.78)	2.48 (1.79)	0.37		

Notes:

*, **, *** denote coefficient significant at 10%, 5% and 1% level respectively (standard errors in parentheses, adjusted by county/fiscal year clusters).

All the regressions in this table include county fixed effects.

Each row represent a regression against LFA and HCD training proportions of the average treatment on the treated effects of welfare entry on employment. These treatment effects were estimated by taking differences of each employment outcome respect to its value on quarter 5 before entry, using WLS regressions with weights for the control individuals given by $(\pi_i/1-\pi_i)$, where π_i is the propensity score for individual i .

Table 8. Differential Treatment Effects of training programs for 1995 Q1 to 1999 Q2 cohorts in 25 CA counties including Welfare Reform indicator and interactions with training probabilities (IV specification)
Outcome: quarterly earnings (\$1,000)

Qtrs After Welfare Entry	Number of Observations	No Training (NT)	LFA-NT		HCD-NT		Welfare Reform		LFA-NT		HCD-NT		F Test	
			Coef	SE	Coef	SE	Coef	SE	Coef	SE	Coef	SE	Welfare Reform	Interact. Coefs=0
1	441	-0.38*** (0.07)	0.35 (0.28)	-0.45 (0.89)	0.01 (0.09)	-0.51* (0.27)	1.25 (1.30)	0.33						
2	441	-0.28*** (0.09)	0.92** (0.35)	-0.80 (1.02)	0.19* (0.10)	-0.94*** (0.32)	1.27 (1.36)	0.01						
3	441	-0.12 (0.09)	1.14*** (0.38)	-1.29 (1.25)	0.26** (0.12)	-1.29*** (0.35)	2.00 (1.56)	0.00						
4	441	0.10 (0.21)	0.04 (1.70)	0.10 (3.68)	0.08 (0.22)	-1.02 (1.23)	3.75 (3.44)	0.53						
5	441	0.25 (0.16)	1.89*** (0.50)	-2.52 (1.72)	0.06 (0.20)	-1.55*** (0.52)	4.29* (2.37)	0.02						
6	441	-0.08 (0.10)	2.69*** (0.47)	-0.65 (1.23)	0.29** (0.14)	-1.91*** (0.53)	1.77 (1.76)	0.00						
7	441	0.19* (0.10)	1.74*** (0.55)	-1.38 (1.44)	0.41** (0.16)	-1.47*** (0.52)	1.58 (2.09)	0.00						
8	441	0.10 (0.19)	2.21*** (0.63)	0.27 (2.82)	0.95*** (0.24)	-1.50* (0.85)	-4.93 (3.64)	0.00						
9	416	0.52*** (0.16)	1.76*** (0.54)	-5.16** (2.40)	0.73*** (0.22)	-2.26*** (0.72)	1.80 (3.31)	0.00						
10	392	0.76*** (0.24)	2.56*** (0.76)	-7.35** (2.96)	0.89*** (0.30)	-4.45*** (1.07)	6.43 (4.63)	0.00						
11	368	0.57*** (0.17)	2.37*** (0.64)	-5.83*** (2.15)	0.35 (0.24)	-4.03*** (1.10)	6.85* (4.10)	0.00						
12	344	0.11 (0.16)	4.06*** (0.76)	-2.05 (2.68)	0.48 (0.43)	-4.60** (1.80)	-0.93 (5.48)	0.00						
13	320	0.54** (0.23)	1.02 (1.04)	-9.75*** (3.67)	0.23 (0.86)	-8.12** (3.67)	20.40** (8.80)	0.03						

Notes:

*, **, *** denote coefficient significant at 10%, 5% and 1% level respectively (standard errors in parentheses, adjusted by county/fiscal year clusters).

All the regressions in this table include county fixed effects.

Each row represent a regression against LFA and HCD training proportions of the average treatment on the treated effects of welfare entry on employment. These treatment effects were estimated by taking differences of each employment outcome respect to its value on quarter 5 before entry, using WLS regressions with weights for the control individuals given by $(\pi_i/1-\pi_i)$, where π_i is the propensity score for individual i .

Figure 1
Training Policies in California
(Average 25 Counties)

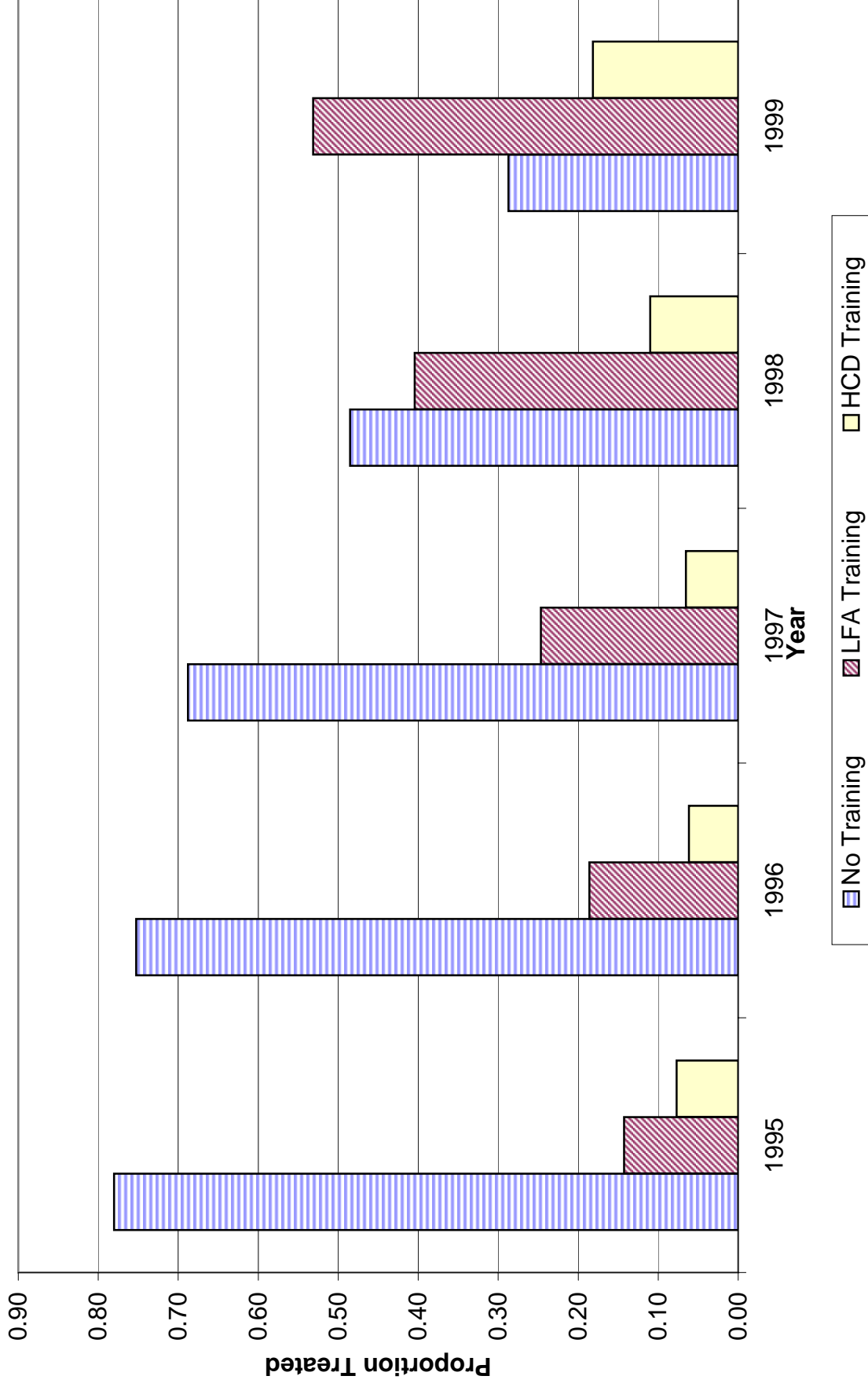


Figure 2
Propensity Score Distribution
Pooled Treated and Control Individuals

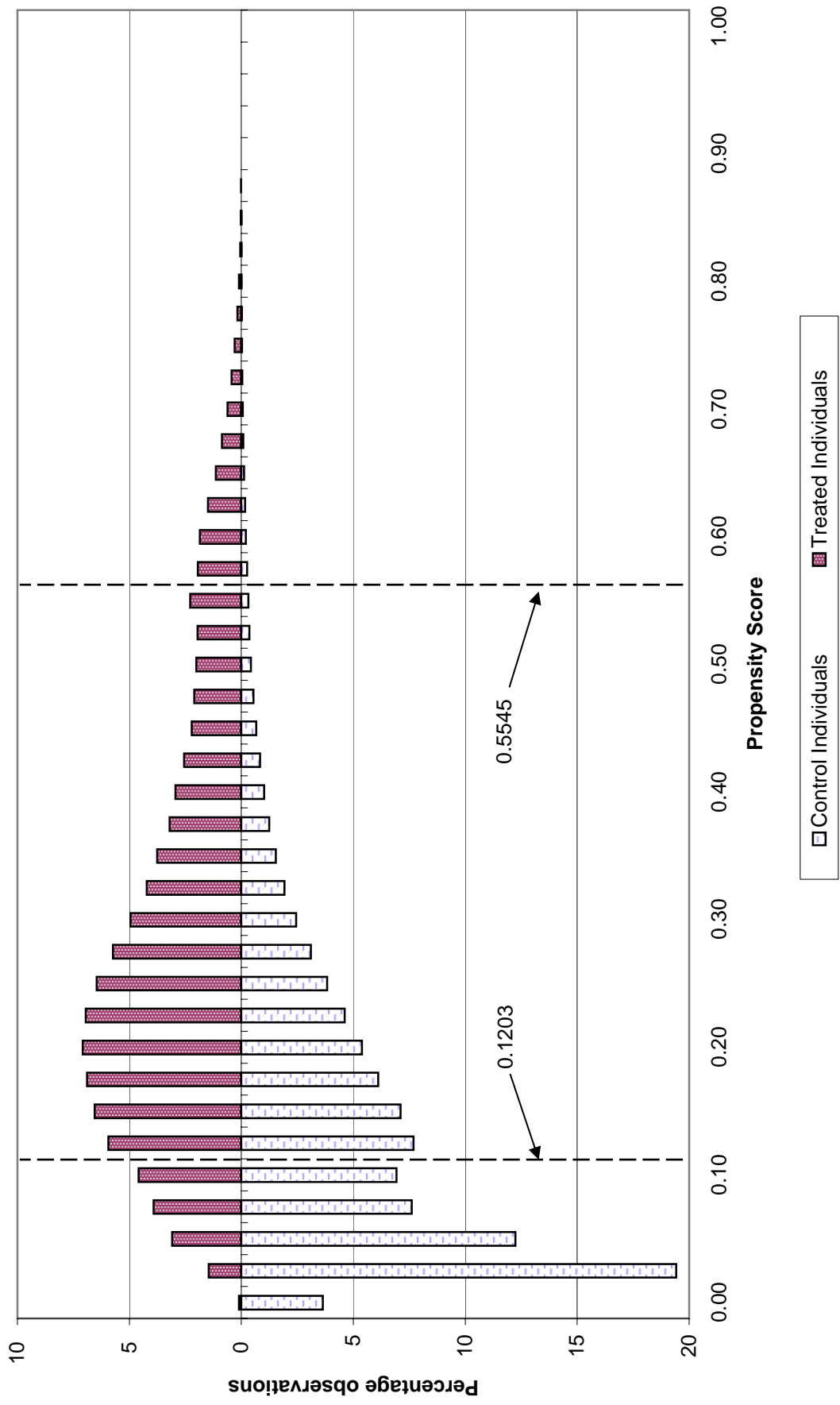


Figure 3
Treatment Effects of Welfare Entry on Employment (levels)
Selected Entry Cohorts (Average 25 CA Counties)

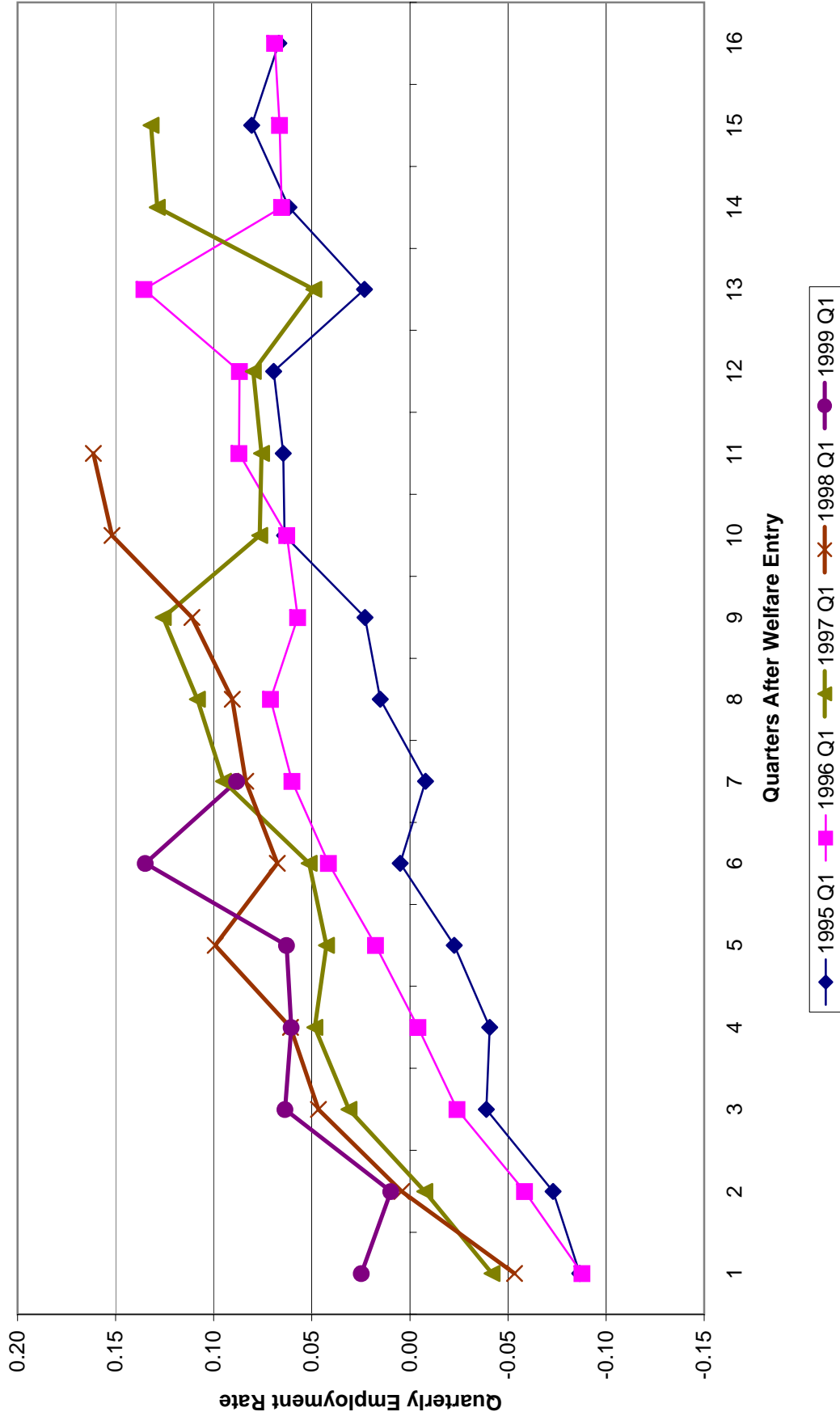


Figure 4
Treatment Effects of Welfare Entry on Employment (diff-in-diff)
Selected Entry Cohorts (Average 25 CA Counties)

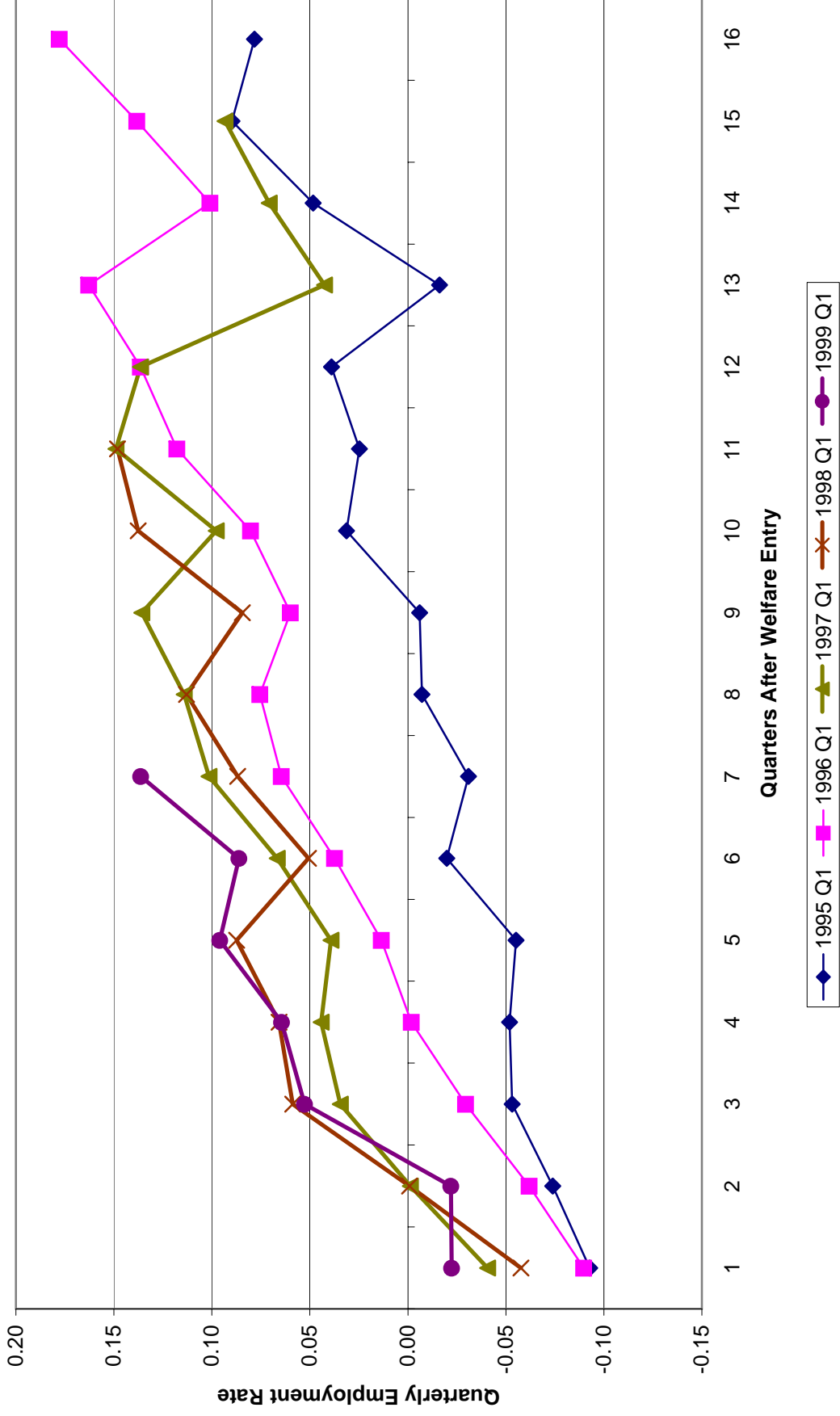


Figure 5
Treatment Effects of Welfare Entry on Earnings (levels)
Selected Entry Cohorts (Average 25 CA Counties)

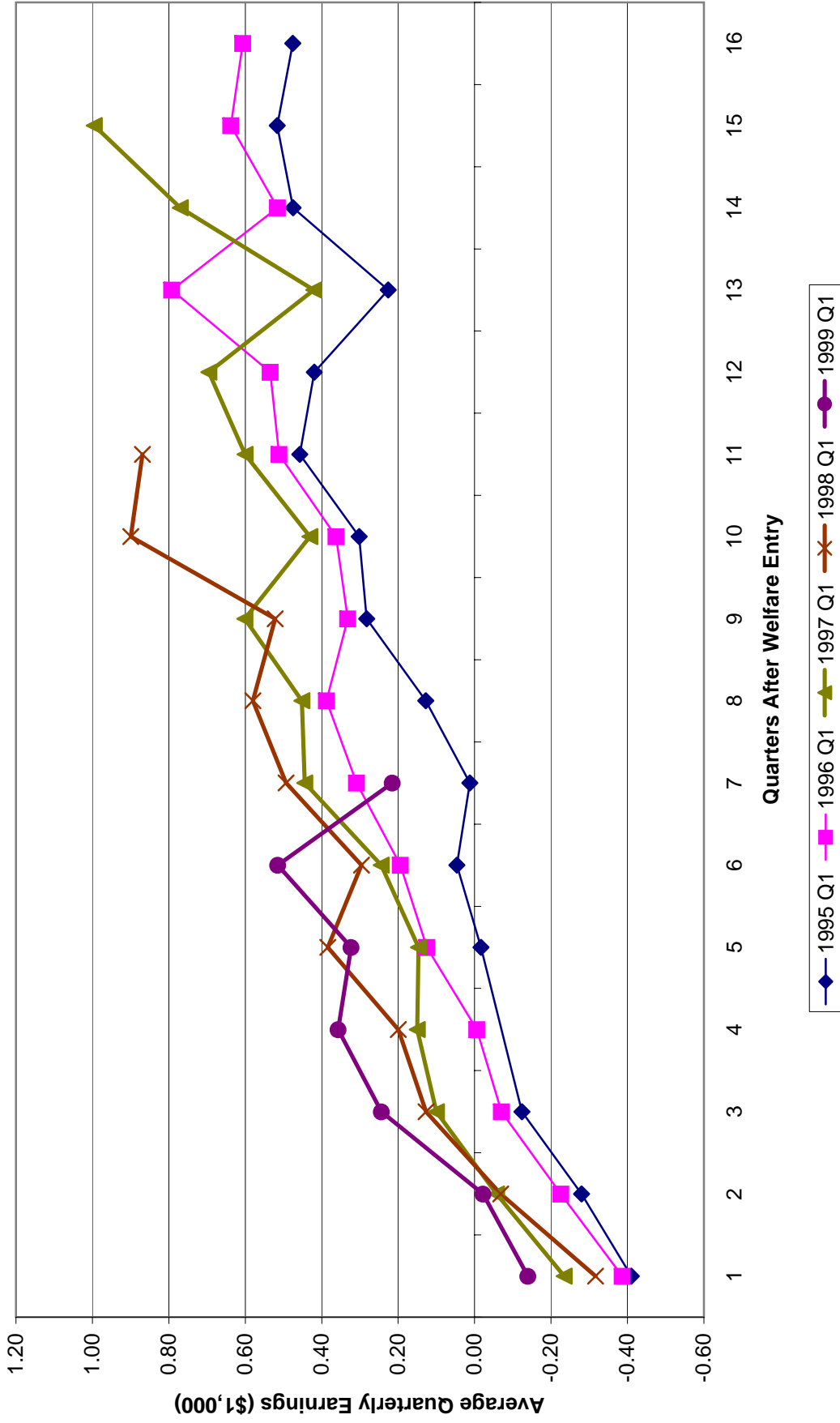


Figure 6
Treatment Effects of Welfare Entry on Earnings (diff-in-diff)
Selected Entry Cohorts (Average 25 CA Counties)

