

IZA DP No. 4024

How Do Training Programs Assign Participants to Training? Characterizing the Assignment Rules of Government Agencies for Welfare-to-Work Programs in California

Oscar A. Mitnik

February 2009

How Do Training Programs Assign Participants to Training? Characterizing the Assignment Rules of Government Agencies for Welfare-to-Work Programs in California

Oscar A. Mitnik

*University of Miami
and IZA*

Discussion Paper No. 4024
February 2009

IZA

P.O. Box 7240
53072 Bonn
Germany

Phone: +49-228-3894-0
Fax: +49-228-3894-180
E-mail: iza@iza.org

Any opinions expressed here are those of the author(s) and not those of IZA. Research published in this series may include views on policy, but the institute itself takes no institutional policy positions.

The Institute for the Study of Labor (IZA) in Bonn is a local and virtual international research center and a place of communication between science, politics and business. IZA is an independent nonprofit organization supported by Deutsche Post Foundation. The center is associated with the University of Bonn and offers a stimulating research environment through its international network, workshops and conferences, data service, project support, research visits and doctoral program. IZA engages in (i) original and internationally competitive research in all fields of labor economics, (ii) development of policy concepts, and (iii) dissemination of research results and concepts to the interested public.

IZA Discussion Papers often represent preliminary work and are circulated to encourage discussion. Citation of such a paper should account for its provisional character. A revised version may be available directly from the author.

ABSTRACT

How Do Training Programs Assign Participants to Training? Characterizing the Assignment Rules of Government Agencies for Welfare-to-Work Programs in California*

A great deal of attention has been paid in the literature to estimating the impacts of training programs. Much less attention has been devoted to how training agencies assign participants to training programs, and to how these allocation decisions vary with agency resources, the initial skill levels of participants and the prevailing labor market conditions. This paper models the training assignment problem faced by welfare agencies, deriving empirical implications regarding aggregate training policies and testing these implications using data from Welfare-to-Work training programs run by California counties during the 1990s. I find that county welfare agencies do not seem to follow a simple returns-maximization model in their training assignment decisions. The results show that, as suggested by political economy models, the local political environment has a strong effect on training policies. In particular, I find that going from a Republican to a Democratic majority in a county's Board of Supervisors has a strong effect on training policies, significantly increasing the proportion of welfare recipients receiving human capital development training.

JEL Classification: C44, D73, I38, J24

Keywords: assignment to training rules, welfare to work programs,
local political environment

Corresponding author:

Oscar A. Mitnik
Department of Economics
University of Miami
P.O. Box 248126
Coral Gables, FL 33124-6550
USA
E-mail: omitnik@miami.edu

* This paper is based on the first chapter of my Ph.D. dissertation. I want to thank V. Joseph Hotz for his tireless encouragement and support, and Guido Imbens for his invaluable guidance. I also would like to thank Sandra Black, Luca Bossi, Moshe Buchinsky, Christopher Cotton, Eduardo Fajnzylber, Jeffrey Grogger, Wes Hartmann, Carolyn Heinrich, Kathleen McGarry, Willard Manning, Enrico Moretti, Federico Najson, Carlos Ponce, Tracy Regan, Jean-Laurent Rosenthal, Duncan Thomas, and participants in several workshops and seminars 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.

1 Introduction

Manpower training programs seek to reduce unemployment and poverty by increasing work-related skills, human capital and the employability of the poor and disadvantaged. For a long time a great deal of attention has been paid in the literature to estimating the impacts, or returns, to these programs. The issue confronting these evaluations, however, is how to address the bias that results from non-random assignment of who gets trained and what type of training they receive. In recent years there has been increasing interest in understanding the processes, or decision rules, governing these choices, in particular as they relate to the roles of governmental agencies, program administrators and caseworkers.

This paper contributes to understanding the assignment decision rules, of government agencies by using data for Welfare-to-Work (WTW) training programs run by counties in California during the 1990s. In 1996, the U.S. welfare system underwent a major reform with the passage of the Personal Responsibility and Work Opportunity Reconciliation Act (PRWORA). One of the key features of the new system was to require that welfare recipients enroll in one or more training or work-related activities, under WTW programs, as a condition of receipt of this temporary assistance.¹ Within these WTW programs, training is typically classified into two types: 1) labor force attachment (LFA) programs that provide welfare recipients with training and assistance in finding jobs; 2) human capital development (HCD) programs that seek to develop basic and work-oriented skills.²

In this paper I seek to understand the decisions made by the agencies that run these programs; in particular, who these programs/agencies train and, more importantly, which type of training they provide to specific participants. I pay special attention to the role played by resource constraints, local economic conditions, and the characteristics of the served individuals. In addition, I analyze the often overlooked role of the political environment of the counties in shaping the programs' decisions.

The literature on the decision process behind training programs has primarily focused on two aspects. First, a recent literature has focused on the role of caseworkers in the assignment to training decisions, and their interaction with statistical profiling or targeting systems. The evidence

¹Note that having a requirement for welfare recipients to participate in training or work-related activities was not new to PRWORA. Since the 1960s, the U.S. welfare system, under the Aid to Families with Dependent Children (AFDC) program, included requirements for participation in employment and training programs by non-exempted welfare recipients. But these provisions were not consistently enforced. The current federal legislation contains stronger mandates and incentives for states to comply with these provisions, especially the participation in work-related activities. Some states even merged the operations of their welfare recipients-related employment and training operation with those for non-welfare recipients established by the Workforce Investment Act (WIA) of 1998, which replaced the Job Training Partnership Act (JTPA). That is not the case for California in the period analyzed, and thus this paper will only address the services provided to welfare recipients.

²For evaluations of the effects of a series of 1990s LFA and HCD programs, see Bloom and Michalopoulos (2001) and Hotz, Imbens and Klerman (2006). Dyke, Heinrich, Mueser, Troske and Jeon (2006) conduct an analysis of more recent WTW programs.

suggests that caseworkers are not very good at attaining the most efficient allocation of potential trainees to types of training (Lechner and Smith, 2007) and that they do not necessarily improve the allocation, even when offered statistical predictions of the best possible treatment for their clients (Behncke, Frölich and Lechner, 2007). This suggests that caseworkers may have their own objectives or may be responding to other objectives in their organizations (and to directives from supervisors). Indeed, Bloom, Hill and Riccio (2003) and Hill (2006) show that the specific details of program implementation and the caseworker job design matter in the allocation decision (and for client outcomes).³

This paper takes a different approach than the studies on the role of caseworkers by focusing on the *overall* decisions made at the county level. The disadvantage of this approach is that it cannot identify the specific role of caseworkers in the decision-making process. However, it has the advantage of taking into account all the factors that affect the allocation decisions, including the role of resource constraints (normally not taken into account in the caseworker studies), the characteristics of the whole cohort of potential trainees (not just those under the responsibility of particular caseworkers), and the influence of the local political environment.

A second strand of the literature concentrates on the modeling and empirical estimation of the effects of incentives and performance standards on the decisions of those public sector organizations that provide training.⁴ The conclusions from this strand of the literature are that performance standards do affect the decisions of program administrators (e.g. Heckman, Heinrich and Smith, 2002) and that those responses are sometimes not intended, with administrators engaging in gaming responses (e.g. Courty and Marschke, 1997 and 2004) and “cream-skimming” (Heckman et al., 2002; Bell and Orr, 2002). During the period I analyze, counties in California were not subject to any binding performance standards.⁵ Thus, I do not study the potential effects of performance standards on assignment to training decisions. However, as Courty, Heinrich and Marschke (2005) note, the practical implementation of performance standards in training programs is a complex principal-agent issue, being hampered by, among other factors, the presence of multiple principals

³A related literature addresses statistical profiling, by which individuals are assigned to programs based on predicted outcomes or program impacts (e.g. Manski, 2000 and 2004; Berger, Black and Smith, 2001; Dehejia, 2005). Profiling mechanisms seek to eliminate discretion in the assignment decisions. As Berger et al. (2001) show, they require a clear definition of the objective function and appropriate design of the profiling mechanism to be effective.

⁴See Dixit (2002) for an overview of the issues surrounding incentives in organizations in the public sector. See Eberts, Hollenbeck and Stone (2002), Hoxby (2002), and Koretz (2002) regarding education policies. Regarding training policies, see Courty and Marschke (1997 and 2004), Dehejia (2005), Heckman, Smith and Taber (1996), Heckman, Heinrich, and Smith (1997 and 2002), and Pepper (2002 and 2003).

⁵Under California’s TANF there was a provision by which certain savings which resulted from the recipients’ exits and improved outcomes would be paid by the state to the counties as performance incentives. This rule was implemented in the last period of this study, fiscal year 1998-1999, when counties claimed a very small fraction of the savings for spending (LAO, 2000). In addition, federal participation requirements were not binding either because caseload reduction “credits” were big enough to make them not binding (see LAO, 2002). A (financial) benefit for California of the participation requirements being effectively not binding was that the amount necessary to satisfy the Maintenance-of-Effort (MOE) requirement introduced by TANF was reduced by \$130 millions per year (LAO, 2000).

and agents. In this context it can still be important to identify the factors that affect the aggregate decisions of local program administrators and agencies in the absence of performance standards, to inform the appropriate implementation of such standards, and the expected results from them.⁶

In addition, this paper makes a contribution to a recent literature on the role of local political variables in the design and implementation of Welfare-to-Work programs. Keiser, Mueser and Choi (2003) and Fording, Schram and Soss (2006, 2007 and 2008) show how the political environment and local political ideology affects the rates of sanctioning of welfare recipients under the TANF program (with higher sanctioning rates associated to a more conservative political environment). Moreover, Fording et al. (2006) rely on cross-county variation in a measure of local political ideology in Florida, to find that the effect of performance standards and performance feedback on sanctioning rates are also influenced by the political environment. They find that more conservative regions react to performance feedback with a much higher sanctioning rate than do liberal regions. This type of effect highlights the relevance of incorporating measures of the local political environment in the analysis of the assignment-to-training decisions.

My analysis begins by modeling welfare agencies as a central planner that assigns workers to different types of training, including a no-training option, so as to maximize the returns to its investments subject to a budget constraint and to the costs of alternative types of training. This “returns-maximization” model of how training programs operate characterizes a “utilitarian” view of the objectives of government agencies (because it implies the maximization of the present value of the participants’ outcomes) and is a natural place to begin to study agency/program decision-making. This is similar to the model of Heckman et al. (2002) who analyze how agencies respond to the imposition of performance standards.

A disadvantage of this aggregate returns-maximization model, is that it does not take into account the principal-agent relationships between managers and caseworkers, nor does it consider the ground-level functioning of the whole system, which could generate a very different set of aggregate preference functions. However, I believe this model serves as a useful benchmark in guiding the empirical exercise. Using this model, I derive testable implications for the training choices of agencies, which take into account the skills-heterogeneity of the participants, the effect of a binding budget constraint on the agency, and the role local market conditions play.

Next, I consider the effects of the political environment at the county level on aggregate training policies. As Fording et al. (2007) point out, the effect of the local political environment may be exerted in several ways, through the actions of local policy makers, administrators and caseworkers,

⁶Starting in 1999, after the period of analysis used in this paper, the federal government implemented a small performance standards program, the High Performance Bonus (HPB) which accounted for less than 1% of TANF’s funding and was to be divided among states that satisfied certain standards (Wiseman, 2004). California received a total of \$123 million due to the HPB, from 1999 to 20001; less than 0.2% of its TANF’s budget in the period. From 2002 to 2005, it received even less, a total of \$34 million (CDSS, 2005). Given the small amounts involved, and the fact that individual counties had no stakes in obtaining HPB awards, it is hard to expect any behavioral responses of individual counties to the HPB.

or advocacy groups. Since the oversight of welfare agencies is conducted by elected officers, a political economy model suggests that public opinion and the beliefs of the median voter should affect the assignment to training policies.⁷ As is well documented in various types of polling data, citizens often have strong views about people on welfare, the appropriateness of having a welfare system and what types of benefits and services people on welfare should receive.⁸ Also, political parties and quasi-public organizations (e.g., unions and special interest groups) often maintain explicit positions, or platforms, regarding the structure of welfare programs.⁹ Lastly, welfare agency administrators and caseworkers may develop their own independent objectives for how their programs are run and which individuals should be served.¹⁰

To capture the local political environment as it relates to the welfare program, I collected a unique dataset with the party affiliation of the members of each county's Board of Supervisors.¹¹ The Board of Supervisors is the body that governs each county in California, applying state law and public policy. It has both legislative powers (it pass ordinances) and executive powers (it has the authority to appoint and remove department heads, and oversees the day-to-day county operations). Therefore, the Boards can determine welfare policies on a county-wide basis. In this study, the party composition of these elected Boards is used as a proxy measure for the *overall* political environment in the county with regard to the welfare program, and not only as a measure of the political or ideological leanings of the supervisors.

As noted above, I conduct my empirical analysis using county-level panel data for the WTW programs in California.¹² California is a well suited state for the application because it runs a very decentralized welfare system in which individual counties have great latitude over which training schemes they offer and how they assign welfare recipients across these programs. Also, as is shown below, there is a great deal of diversity across California's counties in their pool of welfare recipients and their labor market conditions throughout the 1990s.

The empirical analysis focuses on how the proportions of welfare recipients that received three types of treatment—no training, labor force attachment training and human capital development training—vary as a function of the budgets they receive from the State to run the welfare program, the characteristics of their welfare recipients and indicators of the local labor market conditions prevailing in the county. I find that counties determine their training policies based on those factors

⁷There is a debate in the literature on whether politicians respond to public opinion or just to their beliefs and those of their supporters. See Jacobs and Shapiro (2002) and Erikson, MacKuen and Stimson (2002).

⁸See Kluegel and Smith (1986), Cook and Barrett (1992), Bean and Papadakis (1998) and Weaver (2002).

⁹See Kluegel and Smith (1986) and Cook and Barrett (1992).

¹⁰Evidence on the effects of preferences of caseworkers is presented by Heckman et al. (1996). Keiser et al. (2003) and Fording et al. (2006, 2007 and 2008) show how the racial composition of a location, interacted with the political environment can affect the policies towards different racial groups. Also, see Ricucci (2005) for evidence on the impact of the discretionary power of street-level bureaucrats.

¹¹Supervisors are elected in California in non-partisan races, which implies that their party affiliation is not readily available. See Section 4 for details on how the party affiliation information was obtained.

¹²As it is explained in Section 2, during the period analyzed (1994-1999) training for AFDC and TANF recipients in California was offered through the GAIN and WTW programs, respectively.

in ways that appear to be consistent with a simple returns-maximization model. However, when incorporating the effects of local political factors into the empirical model, I find that a simple returns-maximization model is not sufficient to characterize the observed assignment decisions of county welfare agencies. In particular, I find that switching from a Republican to a Democratic majority in the county Board of Supervisors has a strong effect on training policies, significantly increasing the proportion of welfare recipients receiving human capital development training.

These results are important because they suggest that the assignment to treatment decisions reflect a complex interaction between the many agents involved in the determination of training policies and the political environment. This should be taken into account in the implementation and design of performance standards. More importantly, the results raise efficiency concerns about the assignment-to-training policies. Although this has been a contentious issue, there seems to be an emerging consensus in the literature that human capital development programs are preferred in the long-run over labor force attachment programs (e.g. Hotz, Imbens and Klerman, 2006; Dyke, Heinrich, Mueser, Troske and Jeon, 2006). Thus if HCD programs are less favored by counties with Republican Boards, this could have important negative consequences for the long-run outcomes of the welfare recipients living in those areas.

The paper is organized in seven sections. The next section describes the WTW training programs for welfare recipients in California, and shows the evolution over time of the two types of training. The third section presents the optimization problem faced by a welfare office that tries to maximize the expected outcome from its training policies and derives the empirical implications of such a problem. It also discuss the extension of the basic model to allow for effects of local political factors. The fourth section presents the county-level and individual data to be used in the empirical analysis. The fifth section describes the empirical strategy, which entails estimating measures of the distribution of observable and unobservable characteristics of welfare entrants, as a first step, and estimating multinomial logit regressions using the county-level data on training policies in a second step. The sixth section presents the results, while the seventh section concludes.

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 Job Opportunity and Basic Skills (JOBS) program,¹³ 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 was succeeded in 1998 by the Welfare to Work (WTW) program, as part of the California Work Opportunity and

¹³JOBS was intended 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).

Responsibility to Kids (CalWORKs) program, California’s version of the Temporary Assistance for Needy Families (TANF) program.¹⁴ Both programs offered different types of services to welfare recipients who were required to participate (except parents of small children), or face financial sanctions. However, under GAIN counties faced severe funding constraints, and in some counties a large fraction of the caseload remained not served.

Under CalWORKs every adult is required to participate in the WTW program, which implied that counties had to expand their programs to accommodate all the adult caseload.¹⁵ 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 are classified in two groups: 1) those that are work-oriented, which are termed Labor Force Attachment (LFA) programs; and 2) those that are education-oriented which are termed Human Capital Development (HCD) programs. Typically, LFA training is shorter in duration and less expensive to provide than HCD training (see Hamilton et al., 2001 for a discussion of both approaches).¹⁶ For some analyses, HCD training is divided further into Adult Basic Education (ABE) and Vocational Education (VOC), which can be considered as more basic and advanced respectively.¹⁷

Counties in California were given a great degree of freedom in designing their GAIN and WTW programs. This caused 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 cross-county variation from 1994 through 1999 in the proportion of adults receiving the two types of training is presented in Figure 1. Each circle represents a county; their size is proportional to the size of the adult welfare caseload.¹⁸ The 45° line represents the case

¹⁴TANF 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.

¹⁵However, participation in the WTW program does not necessarily imply that a recipient received training services. There are non-training activities that count as participation in the WTW program (e.g., participation in substance abuse help programs).

¹⁶The Labor Force Attachment (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 type of programs are directed to increase the trainees’ human capital by offering basic skills and vocational training programs that are longer in duration.

¹⁷VOC reflects vocational education training activities, while ABE reflects adult basic education, English as a second language and GED-related activities. HCD activities are the sum of VOC and ABE. LFA activities include job club, supervised job search, short- and long-term preparation, alternative work experience, on-the-job-training, subsidized and transitional employment and other job search activities. Unsubsidized employment is not considered a training activity, even though it does count as “participation in a WTW program” for federal participation requirements under TANF.

¹⁸This paper covers the period January 1994-June 1999; after June 1999 a new reporting system for the WTW program (WTW25) was implemented by the state of California which was not fully functional until 2000. Because the comparability of the data from WTW25 with the data from the original reporting system (GAIN25) is far from clear, only the data from GAIN25 is used in this paper.

where the sum of the proportions of both types of training is equal to one, and everybody is trained. Thus, the figure shows an increasing proportion of individuals trained in general over time, but is more marked for LFA training. When the CalWORKs program started (January 1998), there was a larger jump in LFA training than in HCD training. However, in relative terms, the increase was larger in HCD training.

Note that the figure presents data for all the adults on welfare corresponding to a particular period of time. Unfortunately, there is no data available that divides the fraction of recipients between new entrants and non-entrants to welfare. The theoretical model and empirical analysis will be based, however, on the characteristics of new entrants to welfare, because their assignment to treatment is a well defined problem.¹⁹ Moreover, both the GAIN and WTW programs (with more emphasis in the latter) 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 long waiting lists. This is reflected in a “no training” category, which will be explicitly modeled in the next section.²⁰

3 The assignment to treatment decision

In Section 3.1 I present a simple returns-maximization model for the decision making process of a county’s welfare agency that has to assign welfare recipients to different treatment alternatives, and characterize testable implications with respect to the effects of changes in the budget constraint, in local economic conditions and in the initial skills distribution of the welfare entrants. In Section 3.2 I discuss how the basic model can be extended to allow for a role of the local political environment. I suggest an interpretation in which the welfare agency has a richer set of preferences over participants than implied by the simple returns-maximization model, for example across types of training or across individuals with different initial skills. These preferences can arise from the aggregation of preferences of the different stakeholders, for example through the policy and local political process. I show that these preferences can lead to different assignment decisions than those that result from returns-maximization.

3.1 The returns-maximization model

As discussed in the introduction, the centralized view of the decision process sidesteps, by construction, the consideration of any agency issues. To make explicit that the decision maker here is

¹⁹An alternative decision problem, in which the county has to decide the assignment to treatment for a welfare recipient conditional on her (potentially) having already received training before would be much more complex to analyze.

²⁰The “no training” treatment may imply that if the individual stayed long enough on welfare, she could have eventually received some training services, even though that did not happen if the individual was a new entrant.

aggregating all stakeholders in the decision-making process, I will refer to the decision maker simply as the “county.” The hypothesis of the returns-maximization model is that the county’s objective is to maximize the expected returns from its training policies. This seems like a natural starting point, given that it reflects the view that the county’s objective is to obtain the best possible outcomes in the labor market for the welfare recipients: to help them become self-sufficient. Furthermore, this objective function is the same as that assumed by Heckman et al. (2002), Dehejia (2005), Manski (2000, 2004) and Pepper (2002, 2003).²¹

Heckman et al.’s model (2002) is similar to the one below, in particular, because it includes a budget constraint, which other studies do not take into account.²² Additionally, their model allows the decision maker to choose the level of training effort exerted for each individual, which in some sense is similar to the problem in this paper in which the county has to choose among three treatment options that imply different “intensities” of training. The contribution of the model in this paper is to show how the decisions of the county regarding the two training options can be used to empirically test the implications of alternative objectives that govern agencies’ assignment decisions.

Setup Assume a county needs to make training decisions for an entering cohort of welfare recipients. Each individual’s characteristics, which are denoted by a random vector X_i , include her demographic information, educational attainment, previous employment and earnings histories, and, in some cases, evaluations (either objective or subjective) made by county officers on the individual’s potential.²³ I assume that this information is aggregated by the county in a single-dimension index $\theta = \omega(X_i)$, where ω is a known weighting function. Furthermore, I assume that this index is continuous, that it takes on values in the interval $[\underline{\theta}, \bar{\theta}]$, and that it follows a county-specific probability distribution function, $f(\theta)$. I assume that the ability to generate earnings, or to find employment, is positively correlated with θ , which can be interpreted as a measure of an individual’s initial skills level. I will subsequently refer to θ simply as *skills*.

The county has three treatment options for each individual: 1) to not provide any training services (*treatment N*); 2) to provide LFA training services (*treatment L*); or 3) to provide HCD training services (*treatment H*). The county evaluates the effect of each treatment option on an individual with skills θ , by looking at the outcome variable $Y(\theta)$. This variable can represent different outcomes of interest for the county (e.g., earnings, or the probability of finding a job).

Treatments N , L and H have associated, at each period, outcome generating functions $Y^N(\theta)$, $Y^L(\theta)$ and $Y^H(\theta)$, respectively. The outcome functions for treatments L and H can also be rewritten

²¹However, Dehejia also incorporates uncertainty regarding the distribution of outcomes, and allows for a social welfare function that exhibits inequality aversion.

²²Pepper (2003) does consider an optimization model subject to a budget constraint, as one possible source of identification of the potential effects of WTW policies.

²³Some elements of the vector X_i will be observed and some will not be. This issue will be discussed in further detail in Section 5.

as $Y^L(\theta) = Y^N(\theta) + \Delta^L(\theta)$ and $Y^H(\theta) = Y^N(\theta) + \Delta^H(\theta)$, where Δ^L and Δ^H represent the *treatment effects* of training L and H , respectively, and they are assumed to be non-negative functions of θ .²⁴

Because training takes time, the *present value* of the outcome under treatment N could be higher than the present value of the outcome under treatment L , or treatment H for that matter, for some values of θ . I assume that each individual, at the time of treatment, has a future labor life of T periods (fixed). Treatment L takes place over $\tau_L > 0$ periods, while training H takes place over $\tau_H > 0$ periods. Therefore, as HCD training takes longer than LFA training, this implies that $\tau_H > \tau_L$.

The expected present value of the future stream of benefits (in each period t) associated with receiving treatment i is represented by the benefit functions $V^i(\theta) = \sum_{t=\tau_i}^T \rho^t E[Y_t^i(\theta)]$, where ρ is the discount factor, $i = N, L, H$, and $\tau_N \equiv 0$. To simplify the setup, I assume that the expected outcome is constant over time, i.e. $E[Y_t^i(\theta)] = E[Y^i(\theta)]$ for all t , which implies that the benefit functions can be re-expressed as $V^i(\theta) = \kappa_i E[Y^i(\theta)]$, where κ_i is a constant. Without loss of generality it can be assumed that $\rho = 1$, in which case $\kappa_N = T$, $\kappa_L = T - \tau_L$, and $\kappa_H = T - \tau_H$ (and it is easy to see that $\kappa_N > \kappa_L > \kappa_H$).²⁵

Since the cost of training is borne by the county there is no direct cost of receiving training L or training H for the individuals, other than the opportunity cost of attending such training. The cost to the county associated with the provision of treatment N (no training) is normalized to zero, i.e. $c_N \equiv 0$. Thus, the county faces additional direct costs per person c_L and c_H associated with the provision of treatments L and H respectively, where $c_H > c_L$, because HCD training is more expensive than LFA training.²⁶ Finally, the county has a fixed budget B which it must use to cover the expenses associated with providing the training services.²⁷

The county's problem The county seeks to assign individuals to the different treatments such that it maximizes the expected returns to the investment. This problem implies that the county has to assign a treatment for each individual. To make some progress, I impose minimal conditions on the outcome generating functions Y^N , Y^L and Y^H , that make the problem both analytically simpler and, more importantly, informative in terms of the implications associated with the changes in the economic environment faced by the county. Assuming that the outcome generating functions are

²⁴However, note that the differential treatment effect $\Delta^L(\theta) - \Delta^H(\theta)$ can be positive or negative.

²⁵Note that if $\rho < 1$, then $\kappa_N = \frac{1-\rho^{T+1}}{1-\rho}$, $\kappa_L = \frac{\rho^{\tau_L}-\rho^{T+1}}{1-\rho}$, $\kappa_H = \frac{\rho^{\tau_H}-\rho^{T+1}}{1-\rho}$ and it still holds that $\kappa_N > \kappa_L > \kappa_H$.

²⁶Hamilton et al. (2001, pp. 308-309) show the five-year welfare-department portion of the costs of training programs for several counties that participated in the National Evaluation of Welfare-to-Work Strategies (NEWWS). For example, the costs for Riverside County (CA) were \$2,217 and \$3,284 for LFA and HCD training, respectively (in constant 1999 dollars), while for Grand Rapids (MI) the corresponding costs were \$1,610 and \$2,288. However, there were almost no cost differences for the programs in Atlanta (GA), \$3,454 and \$3,466 for LFA and HCD, respectively.

²⁷In the model I abstract from the budget allocation process by which B is available to spend in training programs. This simplification allows me to focus on the determination of training policies, given a budget. The potential endogeneity of the allocation of resources to the training programs will be addressed, however, in the empirical analysis.

concave and satisfy a standard “single crossing” property, which ensures that the benefit functions cross at most once, the decision problem of the county is simplified greatly (the formal assumptions are specified in the Appendix). This implies that the interaction of the functions $V^N(\theta)$, $V^L(\theta) - c_L$ and $V^H(\theta) - c_H$ divide the support of θ , into (at most) three regions, in which all the individuals belonging to a region receive the same treatment.

I denote these three regions over the distribution of θ as R_l (low), R_m (medium) and R_h (high), and use the region subscript to denote the treatment provided in that region (i.e. l, m and h , where each of them can assume values N, L or H). Thus, the county’s problem reduces to choosing two critical values for θ , that define the regions’ limits. Designating θ_l as the critical value that separates regions R_l and R_m , and θ_h as the critical value that separates regions R_m and R_h , the county will choose the θ_l and θ_h that solve the following problem:

$$\begin{aligned} \max_{\{\theta_l, \theta_h\}} W &= \int_{\underline{\theta}}^{\theta_l} [V^l(\theta) - c_l] dF(\theta) + \int_{\theta_l}^{\theta_h} [V^m(\theta) - c_m] dF(\theta) \\ &\quad + \int_{\theta_h}^{\bar{\theta}} [V^h(\theta) - c_h] dF(\theta) - \int_{\underline{\theta}}^{\bar{\theta}} V^N(\theta) dF(\theta) \quad (P1) \\ \text{s.t. } &F(\theta_l)c_l + [F(\theta_h) - F(\theta_l)]c_m + [1 - F(\theta_h)]c_h \leq B, \end{aligned}$$

where c_l , c_m , or c_h will assume the value 0 when valued at treatment N (i.e. $c_N \equiv 0$).

The formulation of (P1) implies that the county only needs to set a decision rule based on the two critical values; assignment to each treatment follows directly from this decision rule. This allows me to study the effects of changes in the environment faced by the county welfare agency simply by analyzing the effects on the critical values for θ .

Solution The first and second order conditions for (P1) are presented in the Appendix. Depending on which type of individual benefits more from each type of treatment, there are six possible interior solutions. All these solutions imply that the support of the skills distribution is divided in three region (low, medium, high skills), and a different treatment is offered to the individuals who fall in each region. Thus, I characterize those six cases by the treatment received by the individuals in the regions of low, medium and high skills, in this order. Denoting them by the notation $[R_l, R_m, R_h]$, where the subscripts refer to the treatment, the cases are:

1. $[R_N, R_L, R_H] \implies$ [No Training, Training L , Training H]
2. $[R_N, R_H, R_L] \implies$ [No Training, Training H , Training L]
3. $[R_L, R_N, R_H] \implies$ [Training L , No Training, Training H]
4. $[R_L, R_H, R_N] \implies$ [Training L , Training H , No Training]
5. $[R_H, R_N, R_L] \implies$ [Training H , No Training, Training L]
6. $[R_H, R_L, R_N] \implies$ [Training H , Training L , No Training].

Budget constraint Here there are two possibilities: If the county has enough funds, such that the budget constraint is not binding, then the welfare agency provides training to all the individuals for which the (net) returns to training are positive. The allocation between training L and training H is determined by the marginal returns to each pair of treatments compared with the marginal costs of treatment. Hence, inside *each* skill region the individuals receive the treatment with the greater (net) benefit compared to the other two. Figure 2 depicts this situation, where panels 2.A to 2.F display Cases 1 to 6 specified above.²⁸

If the budget constraint is binding, it is easy to show that the county will choose the θ_l and θ_h that equalize the ratio of marginal benefits to the ratio of marginal costs associated with the treatments across regions. Hence, the county resorts to substituting between treatments until that condition is satisfied. Note that the county no longer equates just private marginal benefits with private marginal costs, but now has to consider the “social” cost implicit in the fact that providing training to some individuals imply that training has to be denied to other individuals.

To describe exactly how the county attains the optimality condition consider Case 1 as an example. Faced with a binding budget constraint, the county can train fewer people (with respect to the unconstrained case); the issue is how to decide which individuals to train, and which training to offer them. There are two effects. First, because training H is the most expensive treatment, the county chooses to offer fewer people this treatment. As it is shown in Figure 2.A, the individuals “denied” treatment H (that is, individuals who in the unconstrained case would have received training H) will be offered treatment L . But, as more and more individuals around θ_h^u are offered treatment L instead of treatment H , the lost marginal benefits of training increase (as measured by the vertical distance between the curves $V^H(\theta) - c_H$ and $V^L(\theta) - c_L$). Second, if the county denies training to individuals that otherwise would have received treatment L , starting with the individuals around θ_l^u , the lost marginal benefits are also small at the beginning and increase as more and more people are denied training. The new equilibrium critical values are represented in the Figure by θ_l^c and θ_h^c and the shaded areas roughly indicate the lost “social” benefit of substituting for one treatment or the other. Here, note that even if the county has equal concern for every individual (weights them equally), it will choose not to offer every individual her optimal treatment option, because it takes into account the trade-off generated by the resource constraint.

Testable implications of the returns-maximization model The objective of the model is to study how changes in its parameters affect the proportions of people trained (total and in each

²⁸Note that if the individuals do take into account the direct costs of training, c_L and c_H , the solution is the same as would be obtained in the decentralized problem where the individuals are allowed to choose the treatment strategy themselves. For example, individuals could be offered “vouchers”, valid to use in any training of their choice, allowing them to keep any difference between the value of the voucher and the cost of the training. In this way, they would completely internalize the cost of training. As such, the model shares the essential features of the Roy Model of self-selection in the labor market (Roy, 1951), in which heterogeneous agents self-assign themselves to occupations according to a principle of comparative advantage (e.g., Willis, 1986; Heckman and Honore,1990).

type of training). These proportions will be denoted in the remainder of the paper by P_N , P_L , and P_H for treatments N , L and H respectively. Note that $P_N \equiv 1 - P_T$, where $P_T \equiv P_L + P_H$ represents the total proportion of individuals receiving any type of training.

To study the changes in P_N , P_L , and P_H it is clear that the key is to analyze the reaction of the optimal critical values θ_l^c and θ_h^c (where the superscripts will be dropped in the rest of the paper to simplify notation) to changes in the environment faced by the welfare agency.²⁹ Of interest are the effects of changes in the budget B , the effects of changes in a parameter vector Γ that affects the opportunity cost of training $Y^N(\theta; \Gamma)$, and the effect of changes in a parameter vector Ψ that affects the distribution of skills $F(\theta; \Psi)$. In the Appendix I present the formal results, here I just discuss the intuition associated with each.

The discussion above should make clear the testable implications for changes in the budget constraint. If the budget increases, the total proportion of people not trained should decrease, and the total proportion of people receiving training H (the more expensive treatment) should increase, because these individuals are the first affected by the binding budget constraint. The effects on the proportion of people receiving training L are ambiguous, however, depending on whether or not training L was rationed (Cases 3 and 5, positive), or was being used as substitute for training H (Cases 2 and 5, negative) or a combination of both (Cases 1 and 6, ambiguous).

Regarding the effects of changes in the opportunity cost of training, one could concentrate just on Y^N , the potential outcome under no training. If this potential outcome increases, with the training effects held constant, then one would expect that the proportion of individuals trained should decrease (P_N would increase). How P_L and P_H change is more complex to analyze. Given that the budget is fixed and that P_N increases, with everything else held equal, the county has a greater budget available for the people that are going to be trained. Given that training H is always rationed, there should be a substitution of training L for training H (therefore making P_L decrease and P_H increase). However, because training H is the training that takes longer to be completed, the opportunity cost of this training with respect to training N increases more than for training L , making training H less desirable. If this effect is big enough, then the substitution will be from training H to training L , which in turn might even imply (if the savings are big enough, given that training H is more expensive) that the increase in P_L makes P_N actually decrease. The formal result is presented in the Appendix.

If the distribution of skills associated with a given cohort of individuals to be treated changes, it would be useful to understand how the county would change its optimal decisions. This happens to be a particularly difficult issue to explore. In the Appendix I show, for two distributional assumptions (Uniform and Normal), that the effects of changes in the distribution of skills are ambiguous. The intuition is that depending on the (relative) cost of the treatment towards which

²⁹The two critical values are enough to determine the proportions of each treatment: using the general notation for Cases 1 through 6, the proportions are $P_l = F(\theta_l)$, $P_m = F(\theta_h) - F(\theta_l)$, and $P_h = 1 - F(\theta_h)$.

the distribution moves, the county might be able to increase the proportion of people receiving that particular treatment, for the cheaper treatments, but not for the expensive treatments. Therefore, changes that would be easy to analyze if the budget constraint was not binding, become extremely complicated with a binding budget constraint. See the Appendix for details.

3.2 The role of the local political environment

The simple returns-maximization model assumes that only factors associated with returns to training enter the county's decision-making process. However, there are reasons why this might not be the case. As welfare agencies are public bureaucracies there are many ways in which their decisions could reflect other objectives that arise due to the interaction of the objectives and preferences of the involved parties.

First, the general oversight of the welfare agencies is carried out by county supervisors, who are elected officers and determine general policy guidelines, as well as appoint the agency managers. As such, the managers' preferences should reflect the elected officers' preferences. Second, the agency managers and caseworkers might have their own preferences that will affect the agencies' decisions. Third, special interest groups (i.e. advocacy groups) might also influence the decisions of the county welfare agency. Therefore, the combination of these factors might generate an aggregate objective function that is affected by the local political environment. Although it would be interesting to formulate a political economy model of the interactions between different participants in the policy and political process that give rise to the actual training policies pursued by welfare agencies, this is beyond the scope of the current study. Instead, my approach, considering what is feasible from the empirical point of view, is to take the political process as given and assume that it will be reflected in the aggregate objective function of the county (taken as a single unit).

The local political environment could be reflected in different ways in the objective function of the county. For example, one could assume that the county has preferences for a particular type of training. Alternatively, one could assume that the preferences are over particular types of individuals or groups (i.e. based on skills, race/ethnicity, etc.). Empirically, given that only aggregate data is available, it would not be possible to distinguish between these different objective functions. However, empirically, it is still possible to capture whether local political factors influence the counties' assignment to training decisions. This can be used as evidence that the simple returns-maximization model does not properly characterize the aggregate objective functions of the counties, even if the exact alternative objective function is not properly identified.

In what follows I show how one such alternative objective function could affect the assignment to training decisions, namely the case where the county has a preference for a particular type of training. In the Appendix I present an alternative case, where the county has a preference for

individuals in a particular position in the skills distribution.³⁰

In the particular case of LFA and HCD training, it seems reasonable to assume that one of the two types of training could be preferred. The two types of training represent two distinctive approaches with differences that can be deemed philosophical. For example, Hamilton et al. (2001) judge that there has been disagreement over which type of training should be used as the best way of fostering the goal of self-sufficiency. They characterize the differences between LFA and HCD as: “[LFA training] emphasizes quick employment, reflecting the belief that individuals can best build their employability and improve their skills, eventually achieving self-sufficiency, through actual work, even if their initial jobs are minimum wage and without fringe benefits”, while “[HCD training] emphasizes initial investments in short-term education and, in some cases, training, reflecting the view that these investments will eventually enable individuals to obtain higher-wages, longer-lasting jobs with health insurance coverage” (Hamilton et al., 2001, pp. 4). Ideological attitudes towards the role of work and education can also explain the preferences for one type of training or another. In particular, the evidence shows that in the specific case of training policies, and attitude towards welfare reform in general, more liberal individuals favor an education-approach, and more conservative individuals favor a work-first approach (e.g. Nathan and Gais, 1998; Weaver, 2002). In addition, these preferences could reflect heterogeneity in the discount rates associated with the benefits of training by policy makers; it has been shown that LFA training benefits are more immediate, while HCD training benefits appear more in the medium- and long-run (Hotz et al. 2006; Dyke et al., 2006).

Introducing preferences for a particular type of training in the model, considers the case of a county that cares about expected returns from the assignment to training policies, but also has a preference for one type of training over the other. To make the notation simpler, suppose that the shape of the outcome generating functions are such that lower skills individuals do not benefit from training, middle skills individuals benefit from training L , and upper skills individuals benefit from training H (Case 1 in the Section 3.1). Then, denoting by α the weight (preference) given to training L by the county, the problem to solve is as follows:

$$\begin{aligned} \max_{\{\theta_l, \theta_h\}} W &= \alpha \int_{\theta_l}^{\theta_h} [V^L(\theta) - V^N(\theta) - c_m] dF(\theta) \\ &\quad + (1 - \alpha) \int_{\theta_h}^{\bar{\theta}} [V^H(\theta) - V^N(\theta) - c_h] dF(\theta) \quad (P2) \\ s.t. & [F(\theta_h) - F(\theta_l)] c_L + [1 - F(\theta_h)] c_H \leq B. \end{aligned}$$

In the Appendix I show that under “moderate” values of α (i.e. such that both types of training are still offered), the empirical implications regarding changes in the budget and opportunity costs of

³⁰This would arise in a specification where the county’s objective function presents inequality aversion; this type of objective function is proposed by Dehejia (2005). Heckman et al. (2002) also allow for the possibility in their model that there are preferences for a particular group, that is not explained by returns to training.

training of the returns-maximization model still hold. However, the optimal treatment proportions change and are affected by α . The effects of α are clear: higher preference for training L increases P_L and decreases P_H and P_N (see Proposition 3 in the Appendix). If the preferences for training L are extreme (α is “large”), then the testable implications of the returns-maximization model regarding changes in budget and local economics conditions will not hold.

4 Data

The empirical analysis is based on county-level information, and on the aggregation at the county-level of individual data on new welfare entrants. The data on training is constructed as quarterly averages from published county-level monthly reports by every county in California on the number of people participating in the GAIN and WTW programs, and in each of the activities of the program.³¹ Based on these reports, the probabilities of training and participating in activities were constructed.³² As it was mentioned in Section 2, the period under analysis is from 1994 Q1 to 1999 Q2. I use data 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). Table 1 presents statewide and yearly averages of the county-level data that is used in the empirical exercises, while Table 2 presents county and year averages of the characteristics of the sample of entrants that are used in the analysis.

To measure economic conditions at the county-level (as a proxy for the opportunity cost of training), I use published unemployment rates, the employment to population ratio (total and in different sectors), and measures of average earnings for all and specific employment sectors.³³

A county’s budget is approximated by measures of its expenditures. Both the total expenditures and those that measure only the costs associated with providing training services are available. The budget for training is the variable suggested by the theoretical model, but is very likely to be endogenous to the training policies followed by the counties. Therefore, I use the total expenditures associated with running the welfare program (administrative, employment and training services and child care), and do not include the costs of the actual cash aid provided to welfare recipients, as a

³¹As it was 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).

³²The 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. Thus, the proportions of individuals in LFA and HCD training were calculated by summing up the activities in each category, dividing them by the sum the total number of LFA and HCD 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 corresponding month. In a few cases (for small counties) this last ratio produced numbers slightly above one in value. In these cases the ratios were rounded to equal one.

³³The 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.

measure of the “resources” available to the agencies.³⁴

Finally, an important measure is the “political leaning” of the county. The key elected officers, in terms of decisions made at the county-level, are the members of the Board of Supervisors in each county. These Boards have five members each, elected for four years, but with either two or three seats up for election every two years. Each county is divided into five regions, and each region elects one supervisor as its representative in the Board.³⁵

The party affiliation of the members of the Board of Supervisors in each county is then a key measure of the political leaning of the individuals making the actual policy decisions regarding the county welfare program. In addition, how strongly tilted Republican or Democratic is a Board, may reflect the underlying local ideology in the county, which should be correlated with the ideology of the welfare program administrators, caseworkers, and the strength of local advocacy groups.

A significant issue is that *nominally* Supervisor races are non-partisan. In reality, candidates for the Board, in their great majority, do indicate their party affiliation in the campaigns and/or are endorsed by the political parties. However, their party affiliation is not available in the public records of the election results, nor in the county records of the composition of the Boards. To solve this issue I collected a unique dataset with the names of all the members of the Boards of Supervisors in the period under analysis, and matched these names (by county) with the records of Aristotle International, a private company that sells access to voter registration records nationally. The initial matching rate was 60%, but their records contained information that allowed the matching rate to improve to over 70% (for example, by examining whether a particular person voted in the election in which he or she was a candidate). For the 61 (out of 197) Supervisors for which the voting registration records did not allow identifying uniquely party affiliation, a search was done in newspaper databases, counties’ party web-sites and the web-sites of the supervisors to try to identify the party affiliation. Using this method it was possible to identify the party affiliation for all but four supervisors in California.

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 in California (around 90% of the employment of the State).³⁶

³⁴The data on expenditures comes from the County Financial Analysis Bureau at the California Department of Social Services (CDSS). For each quarter the data corresponds to the quarter equivalent of the annual expenditures in the corresponding fiscal year.

³⁵The only exception in the period analyzed was San Francisco County, which has merged the City Council and the Board of Supervisors into one 11-member body. In the 1990s San Francisco’s 11 supervisors were elected in city-wide elections. This changed starting in 2000, with elections being held by region.

³⁶The MEDS (MediCal Eligibility Data System) dataset goes from 1987 to 2000 and contains the welfare use and individual-level information; in addition, it is possible to use case-level data to construct proxies for measures of family characteristics. The Unemployment Insurance (UI) base wage data goes from 1991 to 2000, and comes from

Using the MEDS dataset, new entrants to welfare (defined as individuals that enter welfare for the first time as adults since January 1987) were identified. As mentioned above, only the 25 biggest counties are analyzed; also some individuals were 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.³⁷ An additional sample restriction was that only adults 45 years old or younger are analyzed (because with older welfare recipients, it is not clear how counties perceive the effectiveness of assigning them to training).³⁸

5 Empirical strategy

The implications derived from the theoretical model in Section 3 can be tested empirically, using the county-level data on the total proportions of people trained, and those receiving LFA and HCD training, and the county-level variables presented in the previous section. The following two subsections will detail the econometric specifications used.

5.1 Estimation of skills distribution

As the model emphasizes, it is necessary to control for changes in the skills distribution of each cohort of trainees to be able to compare county data across time and counties. Also, having a measure of the skill distribution permits us to empirically evaluate the effects of changes in this distribution on the counties’ average assignment to training rules.

The distribution of skills is unobservable for each cohort, at least from the point of view of the econometrician. It is very likely that this is true for the county’s welfare agency too, but it is also true that the agency (i.e. the caseworker dealing with a particular individual case) can try to “learn” the value of θ for each individual that enters welfare, by different means. One piece of information is provided by the individual’s previous employment and earnings history, as well as personal information (age, civil status, number of kids. age of kids, etc.) that should be

the quarterly reports filed by employers to the California Employment Development Department (EDD). For 1991 and 1992 this data is available only for a stratified random sample of new entrants (representing approximately 35% of all adult new entrants). Thus, employment and earnings information in years 2 and 3 before entry in Table 2 come from those samples for entrants in 1994 and 1995. This is also why in the individual-level regressions in Appendix Table A1, only 305,120 observations are used.

³⁷Missing 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 five children on welfare.

³⁸In the 1994 Q1 through 1999 Q2 period there were 698,738 welfare entrants in California. Of those, 208,052 (29.8%) were dropped from the analyses because they had missing or invalid information or they were older than 45 years. In addition, 41,050 (8.3%) 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 449,636 new entrants; half of them are concentrated in a 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

very informative regarding the employability and earnings potential of the individual (which is ultimately what θ tries to capture). Thanks to the individual-level data available on all the welfare recipients in California, aggregate measures (at the entry cohort-county-level) of these variables can be constructed and then used in the analyses.

A second piece of information on the underlying θ for the county is provided by the interaction between the county (the caseworker) and the welfare entrant: all counties conduct one-on-one interviews with every entrant to generate an assessment of her potential and weaknesses. In addition, many counties also conduct basic math and language exams to further inform this assessment. Since there is no data on the results of these assessments, this is truly an unobservable from the econometric analysis' point of view (albeit observable for the county). It is also unobservable all other information on the individual that the county observes but is not recorded in the data (most notably educational attainment measures, unfortunately not available in the data sources). Finally, there are some other factors (e.g., the individual's motivation) which are not observed by the county (unless the caseworker is able to gauge them in the personal interviews) or recorded in any data-set.

Let us assume that for an individual i , her skill level is $\theta_i = \omega X_i'$, where X_i is a $1 \times K$ vector of individual characteristics observed by the county, and ω is a $1 \times K$ (known) vector of weights used by the county to aggregate this information. Furthermore, assume that X_i is partitioned into two vectors, $X_i = [X_i^o, X_i^u]$, where X_i^o is the vector of characteristics observed in the econometric analysis, and X_i^u is the vector of characteristics that are unobserved in the econometric analysis. Similarly ω can be partitioned as $\omega = [\omega^o, \omega^u]$.

Because the regressions of interest are at the county-level, it is necessary to introduce measures of the characteristics of the distribution of θ ; for example by including in the county-level regressions cohort/county means of the observed variables X^o . However, this has two limitations. First, the coefficients associated with these measures do not have a clear interpretation. Second, this strategy leaves out X^u which might include very important factors in the county's decision. Arguably, X^o includes factors like employment and earnings histories that are considered as crucial in the selection into training decision, by the training literature³⁹; but other important factors used by the county to make decisions, may be left out.

My strategy to address this problem is to use the individual-level information available for welfare entrants to estimate an index for the left-out individual components, and then include features of its distribution in the aggregate-level regressions. Specifically, in the tradition of the estimation of earnings equations literature, I calculate earnings equations for the welfare entrants in a period *prior* to entry, and interpret the residuals from these regressions as measures of the unobservable component of the skills of the individuals.

³⁹The whole "selection on observables" approach to the non-experimental literature of estimation of training effects, is based on the assumption that a rich set of variables is available on the individuals, with earnings and employment histories playing a central role.

It is crucial then to obtain the best possible prediction of earnings. Given that a large proportion of the individuals were not working before they entered welfare, there is a substantial mass at zero earnings, and it is important to estimate the earnings equations in a way that explicitly deals with this mass at zero

There are two usual approaches in the literature. One is to use a Heckman selection model where the decision to participate in the labor force is estimated jointly with the log(earnings) equation. The second is to use what is called a “two-part” model, which models, in a first step, the labor force participation decision, and in a second step the earnings-generating equation, conditional on participation. There has been a debate in the literature over the merits of the two approaches, in particular regarding prediction of $E[Y|X]$, which is of main interest in this case (see Manning, Duan and Rogers, 1987; Leung and Yu, 1996). The evidence weakly favors the two-part model over the Heckman selection model on this basis (Leung and Yu, 1996). One of the issues is how well the Heckman selection model deals with a particularly high mass at zero, as it is in this case; over 50% of welfare entrants are not working before entry. There are two possible advantages in that regard of the two-part model. First, it allows for distributional assumptions that are more flexible than the Normal assumption; second, it allows for the estimation of the positive portion of the earnings distribution in levels without having to take logs, and then convert back to levels to obtain the predicted \hat{Y} for each individual (see Mullahy, 1998; Blough, Madden and Hornbrook, 1999; Manning and Mullahy, 2001).

Here, I estimate both a standard Heckman selection model, and a two-part model, to compare the results. In the case of the Heckman selection model, the earnings equation is estimated in logs, and the participation equation by a probit model. Then, for each individual $\hat{Y}_i = \Pr(\widehat{Y}_i > 0) \exp(X_i \hat{\beta} + .5\hat{\sigma})$, where $\hat{\beta}$ and $\hat{\sigma}$ are estimated coefficients from the log(earnings) equation (Manning et al., 1987), and homoskedasticity is maintained.

For the two-part model, first the probability of earnings being positive is estimated with a logit model; and second the earnings equation (for positive earnings only, in levels) is estimated with a GLM model using a Gamma function and log link function. Then, for each individual, $\hat{Y}_i = \Pr(\widehat{Y}_i > 0) \exp(X_i \hat{\delta})$, where $\hat{\delta}$ is the estimated coefficient vector from the GLM estimation. The main advantage of this two-part model specification is that it uses a Gamma function which allows for a big mass at earnings close to zero, and it can be estimated under heteroskedasticity. In both cases, after \hat{Y}_i has been obtained, we can calculate the residual $u_i = Y_i - \hat{Y}_i$, which will be the measure of the unobservable of the individual i .

Appendix Table A1 shows the results from estimating these two models. In the two-part model the same variables were included in both stages of the estimation, while as it is usual for the Heckman selection model, some exclusion restrictions were imposed (the excluded variables were a dummy for a single parent family, a dummy for whether the individual received AFDC as a child and four family-composition variables). In addition to the typical demographic and family

variables, the models include employment history indicators and measures of local labor market conditions.

Since the measure of unobservables is a proxy for relatively “permanent” components in the earnings generating function of each individual, I use as the dependent variable the sum of the earnings observed for quarters 5 to 8 before first time of entry into welfare. Using, for example, quarters 1 to 4 before entry would have implied capturing part of the income shocks that explain entry into welfare in the first place (Ashenfelter’s dip). Alternative specifications with different quarters before entry were tried, and as long as quarters 1 to 3 were not used, all the other cases showed similar results. In addition, using four quarters together slightly diminished the issue of zero earnings (compared to using single quarter measures). Note that the estimation was conducted for all the cohorts of welfare entrants across all counties, which were pooled together, and including county and quarter dummies as a way of obtaining a measure that is comparable across all counties and time periods.

After the individual residuals are obtained, we investigate whether the estimated distributions differ across the two specifications. A Spearman rank correlation coefficient was estimated to be 0.902 between the two distributions, which suggests that they are both essentially capturing the same features. Figure 3 makes the comparison more clear by plotting the relative position of each individual, by deciles of the unobservables distribution, according to each method. The size of the circles indicate the number of individuals in each point. The vast majority of the individuals are in the 45° line, implying that both models classify them in the same decile, or are at most a decile off. For a small number of individuals, the two-part model tends to allocate them in higher deciles than for the Heckman selection model (all those above the 45° line), while the opposite is true for an even smaller number of observations. This is probably a reflection of the two-part model being able to better handle the mass of individuals at low earnings. In the next section, the unobservables estimated from the two-part model will be used, although the results are not very sensitive if the unobservables estimated with the Heckman selection model are instead used.

To understand the changes in the unobservables over time, Figure 4 presents the CDF of the unobservables (from the two-part model) for the initial and final period in the analysis. The unobservables are standardized to have a mean of zero and a standard deviation of one, and the percentiles displayed in the figure are the ones corresponding to the pooled distribution. The figure shows that from 1994 to 1999 a compression of the distribution of unobservables has occurred; the percentage of individuals in 1999 that are in the 10th and 90th percentile of the pooled distribution has decreased (this is true also for the 25th and 75th percentile). The question of interest here, of course, is how did counties change their assignment to training decision in view of these movements in the unobservables distribution?

5.2 County level regressions

The model presented in Section 3, provides testable implications on the effects of several variables on the proportions of the different treatments, P_N , P_L , and P_H . This suggests estimating the following three equations:

$$P_{N,ct} = \beta_{0N} + \beta_{1N}B_{ct} + \beta_{2N}E_{ct} + \beta_{3N}M_{\hat{U},ct} + \beta_{4N}\bar{X}_{ct} + \beta_{5N}Z_{ct} + v_{N,ct} \quad (1)$$

$$P_{L,ct} = \beta_{0L} + \beta_{1L}B_{ct} + \beta_{2L}E_{ct} + \beta_{3L}M_{\hat{U},ct} + \beta_{4L}\bar{X}_{ct} + \beta_{5L}Z_{ct} + v_{L,ct} \quad (2)$$

$$P_{H,ct} = \beta_{0H} + \beta_{1H}B_{ct} + \beta_{2H}E_{ct} + \beta_{3H}M_{\hat{U},ct} + \beta_{4H}\bar{X}_{ct} + \beta_{5H}Z_{ct} + v_{H,ct} \quad (3)$$

where the subscripts c and t refer to county and time; $P_{N,ct}$, $P_{L,ct}$ and $P_{H,ct}$ are respectively the total proportion of people not trained, the proportion of people that received LFA training, and the proportion of people that received HCD training; B_{ct} represents the budget; E_{ct} represents measures of local economic conditions; $M_{\hat{U},ct}$ represents measures of the estimated unobservables distribution for welfare entrants in county c at time t ; \bar{X}_{ct} represents the average observed characteristics of the welfare entrants in county c at time t ; and Z_{ct} represents measures of the local political environment. In addition county and time dummies are included in the equations, as a way to control for invariant county- and time-specific factors.

One option to estimate these equations would be to run linear regressions on the proportions of the three treatments. However, this has the unattractive feature that the estimated proportions could lay outside the (0,1) range. A better alternative is to estimate a multinomial logit model for grouped data, where only two out of three equations are actually estimated; after the estimation, marginal effects can be calculated for the three proportions.

An important issue in the estimation is choosing the appropriate measure of resources available for training. The variable “training budget” (see Table 2) is derived from the report of expenditures of the counties on the training programs. Therefore, it is very likely that this variable will be endogenous to the training policies of the counties. To deal with this issue I use instead the total budget (expenditures) of the county. This seems to be a reasonable approximation for resource availability, which should not be affected by endogeneity concerns with respect to the training policies. In the period analyzed, the total county-level budget allocations to counties were decided at the state-level by the California Department of Social Services, and therefore are exogenous to the counties’ training policies.⁴⁰

A second issue concerns to the (potential) endogeneity of the characteristics of new welfare

⁴⁰For the 1994-1999 period analyzed in this paper, two different budget allocation systems were in place in California. From 1994 to 1997, the State essentially decided how much money would be allocated to different programs in each county, generating a situation where “county budgets reflect state policy choices” (LAO, 1996). From 1997 to 1999, budget allocation for employment and child care services, but not for administrative costs, was based on a “statewide model” based on historical spending patterns (LAO, 2001). Under neither regimes, were budget allocations based on the training policies in the county. Starting in 2000, however, the budget allocation process changed to better reflect the county policy choices (LAO, 2002).

entrants and training policies. There is evidence that individuals do respond strategically to changes in welfare regulations (e.g., Grogger and Michalopoulos, 2003). The empirical specification controls for measures of the unobservables distribution of welfare entrants, and includes county and time dummies, which should allay concerns that the results are driven by unobservable heterogeneity across counties or time. In addition, it is hard to argue that new entrants will apply to welfare just because of the availability of training services; similar services are easily available outside the welfare system, for example through the Workforce Investment Act (WIA) program, and the cash component of welfare is likely to be the driving force in entry decisions. What would pose a problem, however, is if counties actively used their training policies to discourage welfare entry. There is evidence (e.g. Fording et al., 2006) that this type of behavior arises in response to welfare offices failing (or being at risk of failing) to satisfy minimum standards, under a performance standards regime. However, during the period analyzed in California, as it was explained in footnote 5, there were no performance pressures on the counties, due to the rapid decrease in the welfare caseload in the period. Thus, it seems reasonable to assume that reverse causality of training policies on new entrants characteristics is not present here.

6 Analysis of results

Table 3 presents the marginal effects, evaluated at the means of the covariates, from estimating multinomial logit regressions on the proportions of the three treatments of interest (no training, LFA training, HCD training).⁴¹ The complete model estimated is presented in Appendix Table A2, while Table 3 presents the variables of interest. All the specifications include controls for average demographic characteristics, and county and fiscal year dummies; the standard errors are robust to heteroskedasticity, and corrected by county/fiscal year clusters.

Effects of changes in the unobservables distribution and budget The characteristics of the individuals, both observables and unobservables play a role in the determination of the training policies.⁴² The proportion of individuals in each cohort (i.e. county/quarter) that are in the bottom and upper deciles of the *pooled* unobservables distribution are included as a way of characterizing the changes in the skill distribution.⁴³ The proportions are expressed as deviation with respect to the “expected” proportion of 0.10, which should facilitate the interpretation. Models (1) and (2) make it clear that the unobservables matter in different ways for different types of training (the models differ only in the inclusion of an indicator variable for periods after welfare reform).

⁴¹When estimating marginal effects, if the variable is a dummy, the marginal effects represent the change from 0 to 1 in that variable.

⁴²The coefficients associated to the demographic characteristics of the entrant cohorts are shown in the Appendix Tables.

⁴³Other deciles were included but did not appear as statistically significant. Note that the definition of deciles is not cohort-specific, but rather from the pooled model used to estimate the unobservables.

An increase in the proportion of individuals in the lower decile decreases the proportion of LFA training, and increases that of HCD training (even though the latter effect is not statistically significant), and an increase in the upper decile decreases both types of training. As a reference, to better interpret the results, consider the effects of the changes in the unobservables distribution depicted in Figure 4. The proportion of individuals in the upper decile decreased from 0.111 in 1994 to 0.088 in 1999, for a net change of -0.023. The coefficient for the 90th percentile for LFA in Model 2 (-0.833) implies that for every 0.01 decrease in the proportion of individuals in the 90th percentile, LFA training will increase by 0.0083. Thus, the observed changes in the upper part of the unobservables distributions would explain an increase of $-0.023 \times -0.833 = 0.019$ in the proportion of LFA training (or close to 5% of the total change in the period from 0.11 to 0.53). The same type of calculation for HCD training shows that the compression of the upper part of the unobservables distribution accounts for an increase of 0.013 (around 14% of the change in HCD training in the period). Referring to the six possible cases presented in Figure 2, and to the testable implications from the returns-maximization model, the results from Model (1) and (2) suggest that Case 6 is the empirically relevant one, with HCD training being offered to the lower portion of the distribution, while LFA is offered to the middle part and no training is offered to the upper portion of the skills distribution.

The effects of an increased budget, as measured by per capita expenditures, also goes in the direction predicted by the model; it increases HCD training and decreases no training. The expected sign of the effect on LFA training was ambiguous; the empirical result is positive. Note that the average total budget increases modestly in the period (around 10% from 1994 to 1999), but because of the large decrease in caseload, the per capita resources almost doubled. The results in Model (2) suggest that those changes explain around half of the overall change in LFA training and around two-thirds of the change in HCD training in the period.

Model (3) interacts the expenditures variable with the unobservables measures, with the results showing an even larger effect of the changes in unobservables in the upper part of the distribution, when the budget is higher.

Effects of local economic conditions To capture local economic conditions, several variables were considered. Unemployment rates appear as statistically significant, as well as employment and earnings in the service sector, which have been used previously to capture county-level labor market conditions (e.g., Hoynes, 2000). Retail sector variables showed similar effects, which should not come as a surprise given that retail is one of the more important sectors for low wage workers, and therefore is the sector where welfare recipients would most likely try to obtain jobs. However, they were dropped due to high correlation with the service sector variables.

Interestingly, increases in unemployment rates have a marginally statistically significant negative effect on the proportion of HCD training, which is not consistent with the idea that when the outside

option gets worse, the implicit time cost of HCD training decreases. The increase in LFA training associated with higher employment rates is intuitive: when the markets are demanding workers, it is easier to implement the LFA approach.

Effects of the local political environment The results in Table 3 appear consistent with the testable implications of the returns-maximization model. However, as it was discussed in Section 3.2, they may also be compatible with models in which the welfare agency administrators exhibit moderate preferences for a particular type of training or moderate inequality aversion. Adding measures of the local political environment could capture these types of preferences.

Table 4 shows the marginal effects from regressions similar to the ones presented in Table 3 (see the regressions in Appendix Table A3), but adds information on the party composition of the county Boards of Supervisors. Two different measures are considered. The first one just measures if the Board has a Democratic majority (at least 3 out of 5 members). To understand the variation captured with this variable, Figure 5 presents all the counties in the study, every year (weighted by size), according to whether the Board had a Democratic or Republican majority in 1994. Changes in the following years from the left to the right of each graph (or vice versa) signify a Board that has changed the majority party. Around 30% of the counties/years show this type of changes, providing a key source of exogenous variation to identify the effects of changes in the local political environment. In addition, Figure 6 represents the same data as in Figure 1, but now classified by whether the Board has a Democratic or Republican majority. It is clear from the picture, that the increase across the board in training has not occurred in the same way for all counties, with a clear trend of Democratic counties moving towards a higher proportion of HCD training (compared to Republican counties), while increasing LFA training at the same time.

This is confirmed by the regressions in Table 4, Models (1) to (3), which show that a change of a Board from Republican to Democratic increases the proportion of HCD training by around 0.03. This is a large effect, considering that the average proportion of HCD training for the period is 0.09. The effects of budget and local economic conditions are not affected by the inclusion of this political variable. Regarding the effects of changes in the unobservables distribution, Models (1) and (2) suggest that once the political environment is taken into account, changes in the upper part of the unobservables distribution are the only ones that matter (in the same direction as before). Model (3) interacts the political variable with the unobservables variables. The results for the interaction of a Democratic Board with the 10th percentile are very interesting; the negative sign for no training and positive sign for LFA training (both significant) suggest that the compression in the lower part of the unobservables distribution increases no training and reduces LFA training. However, it is puzzling that this is not reflected in an (marginal) increase in HCD training.

Finally, Model (4) in Table 4 considers the case where the Democratic majority variable is divided in two: ‘weak’ and ‘strong’ majority (3 out of 5 supervisors and 4 or 5 out of 5 respectively).

Interestingly, going from a ‘weak’ Democratic majority to a Republican majority generates a similar reduction in HCD training as before (0.028). However, if a ‘strong’ Democratic county switches to Republican, then the negative effect on HCD is more than doubled, to 0.062. This is consistent with a political economy model where a strong majority can push policies in their preferred direction to a larger extent than a weak majority.

A finer classification of HCD programs One of the issues raised by the recent literature evaluating the effects of WTW programs is the need to better account for the heterogeneity of actual services received within broad training categories (e.g., Dyke et al., 2006). To explore this issue, I divided the HCD types of training into its two components: 1) adult basic education (ABE), which include ESL and GED courses, as well as adult basic education; and 2) vocational education training (VOC), which includes courses at community colleges and oriented to a trade profession.

Table 5 presents the marginal effects from running a multinomial logit model with four categories, instead of three, for the two sets of political variables explored (the regression results are presented in Appendix Table A4). Analyzing Model (2) it is interesting to note that the variable Democratic majority seems to affect almost equally the proportions of ABE and VOC training. On the other hand, expenditures and local economic conditions affect only VOC, not ABE. The effects of changes in the upper part of the unobservables distribution are similar as before; in the lower part, however, it appears to have an interesting positive effect on ABE, but not on VOC. This is consistent with the ABE services being offered to the individuals with the lowest level of initial skills, and confirms that Case 6 of the model in Section 3 is probably an adequate representation of the types of individuals receiving each type of treatment. Finally, the same basic pattern of results emerge from Model (3), when considering ‘weak’ and ‘strong’ Democratic majorities. For both types of majorities, there is a large and significant effect on the proportion of VOC training; for ABE training, however, the effect is significant only for a ‘weak’ majority.

Summary It is clear that the political environment of the county, as measured by the party composition of the Board of Supervisors, has a large positive effect on the proportion of welfare recipients that receive HCD training. The same is true for the components of HCD training, adult basic education and vocational education. Moreover, the stronger the Democratic majority in a board of supervisors, the larger is the positive effect on those types of training. In addition, the results associated to the measures of the unobservables distribution allow identifying the type of individuals (as characterized by unobservables) that receive alternative treatments. The results suggest that lower-skilled (unobservables) individuals are offered HCD training, while middle-skilled individuals are offered LFA training. The highest-skilled individuals are those not offered training. This is probably consistent with a “work-first” approach where the objective is to offer the least amount of training possible, such that individuals can obtain a job as quickly as possible.

7 Conclusion

I model the training assignment problem faced by county welfare agencies and consider alternative assumptions about the objective function of the agencies when assigning welfare participants to different types of training. I derive empirical implications regarding aggregate training policies, and test them using data for Welfare-to-Work training programs run by California's counties during the 1990s. Overall, my findings are that counties determine their training policies based on resource constraints, labor market conditions, and the characteristics of the individuals to be served in ways that appear to be consistent with a simple returns-maximization model. However, when incorporating the effects of local political factors, I find that a simple returns-maximization model is not sufficient to characterize the observed assignment decisions of county welfare agencies. In particular, I find that switching from a Republican to a Democratic majority in the county Board of Supervisors has a strong effect on training policies, significantly increasing the proportion of welfare recipients receiving human capital development training.

These results are important for two reasons. First, they underscore the need to properly take into account the complex interaction between the many agents involved in the determination of training policies and the political environment. The period analyzed was mostly devoid of performance standards pressure, but the most recent policy emphasis has been on imposing performance standards on Welfare-to-Work programs. As Fording et al. (2006) show, differences in the local political environment can affect the way in which performance standards systems operate. Thus, identifying and recognizing the presence of local political factors when devising performance standard systems may be of utmost importance to attain the desired performance objectives.

Second, and more importantly, the results raise efficiency concerns about the assignment-to-training policies. There seems to be an emerging consensus in the literature that human capital development programs are preferred in the long-run over labor force attachment programs (e.g. Hotz et al., 2006; Dyke et al., 2006). Thus, if HCD programs are less favored by Republican-leaning counties, this could have important negative consequences for the long-run outcomes of the welfare recipients living in those areas. Given these potential efficiency costs, future research should attempt to understand the mechanisms behind the political process that gives rise to the assignment to training decisions, with the objective of finding ways of minimizing efficiency losses.

Appendix

Assumptions to obtain (P1) The following are the assumptions necessary to obtain (P1).

Assumption 1

The outcome generating functions $Y^i(\theta)$ ($i = N, L, H$) are:

- i) twice differentiable
- ii) concave: $Y_{\theta\theta}^i \leq 0$
- iii) strictly increasing: $Y_{\theta}^i > 0$ for all $\theta \in [\underline{\theta}, \bar{\theta}]$
- iv) bounded from below and above: $Y^i(\underline{\theta}) \geq 0$ and $Y^i(\bar{\theta}) < K$, where the constant $K < \infty$.

Assumption 1 implies that there are always positive but decreasing returns to skills for any treatment and that the problem will be well specified. Note that with Y^N concave, it is sufficient that the treatment effects Δ^L and Δ^H are concave, to have outcome functions Y^L and Y^H that are concave.⁴⁴

Assumption 2

There is an ordering of the (present value of) expected marginal returns to skills for the three outcome functions Y^N , Y^L and Y^H that is unchanged for all $\theta \in [\underline{\theta}, \bar{\theta}]$. That is, for all triplets of outcome functions (Y^i, Y^j, Y^k) ordered by their marginal returns, it is satisfied that

$$\kappa_i Y_{\theta}^i < \kappa_j Y_{\theta}^j < \kappa_k Y_{\theta}^k \quad \forall \theta, \quad i, j, k = N, L, H, \quad i \neq j \neq k.$$

Assumption 2 has a clear economic interpretation: if a particular treatment is better for individuals of higher skills, then the marginal returns to skills are also higher for these individuals. This is a standard single crossing property, and is crucial to be able to simplify the decision problem of the county.⁴⁵

Assumption 3

The cumulative distribution function of θ is differentiable and strictly increasing in θ : $F_{\theta}(\theta) \equiv f(\theta) > 0$, for $\theta \in [\underline{\theta}, \bar{\theta}]$.

Assumption 3 guarantees that there are no regions of θ in which the density function is not defined, and it is necessary to assure that the first order conditions of the problem are well defined.

It is straightforward to prove that under Assumptions 1 and 3 a solution always exists.⁴⁶

First order conditions for (P1) The first order necessary conditions for (P1) (using a Lagrange multiplier λ for the budget constraint) are the following:

$$\begin{aligned} \theta_l : [V^l(\theta_l) - c_l]f(\theta_l) - [V^m(\theta_l) - c_m]f(\theta_l) + \lambda f(\theta_l)(c_m - c_l) &\leq 0 \\ \theta_h : [V^m(\theta_h) - c_m]f(\theta_h) - [V^h(\theta_h) - c_h]f(\theta_h) + \lambda f(\theta_h)(c_h - c_m) &\leq 0 \\ \lambda : B + F(\theta_l)(c_m - c_l) + F(\theta_h)(c_h - c_m) - c_h &\geq 0. \end{aligned} \tag{A1}$$

When the budget constraint is binding, it is easy to show that the county will choose the θ_l and θ_h that satisfy the following condition (where the superindexes stands for “constrained”):

$$\frac{V^m(\theta_l^c) - V^l(\theta_l^c)}{c_m - c_l} = \frac{V^h(\theta_h^c) - V^m(\theta_h^c)}{c_h - c_m}. \tag{A2}$$

That is, the treatment regions are chosen in a way that equalizes the ratio of marginal benefits to the ratio of marginal costs of the treatments across regions.

⁴⁴ Although this is not necessary as long as $Y_{\theta}^N > \Delta_{\theta}^L$, $Y_{\theta}^N > \Delta_{\theta}^H$ and $|Y_{\theta\theta}^N| > |\Delta_{\theta\theta}^L|$, $|Y_{\theta\theta}^N| > |\Delta_{\theta\theta}^H|$.

⁴⁵ This condition is analog to what is known in the incentives literature as “Spence-Mirrlees” condition or “constant sign” condition (see for example Salanié, 1997, pp. 31).

⁴⁶ Assumptions 1 and 3 guarantee that the objective function W is continuous and that the budget set is compact.

Second order conditions for (P1) The second order conditions that need to be satisfied around a maximum for P1, denoting by L the Lagrangian, are:

$$\det(M) \equiv \det \begin{bmatrix} 0 & \frac{\partial^2 \mathcal{L}}{\partial \theta_l \partial \lambda} & \frac{\partial^2 \mathcal{L}}{\partial \theta_h \partial \lambda} \\ \frac{\partial^2 \mathcal{L}}{\partial \theta_l \partial \lambda} & \frac{\partial^2 \mathcal{L}}{\partial \theta_l^2} & \frac{\partial^2 \mathcal{L}}{\partial \theta_l \partial \theta_h} \\ \frac{\partial^2 \mathcal{L}}{\partial \theta_h \partial \lambda} & \frac{\partial^2 \mathcal{L}}{\partial \theta_l \partial \theta_h} & \frac{\partial^2 \mathcal{L}}{\partial \theta_h^2} \end{bmatrix} > 0 \quad (\text{A3})$$

where (valued at the critical values θ_l^* and θ_h^* that solve (P1)):

$$\begin{aligned} \frac{\partial^2 \mathcal{L}}{\partial \theta_l^2} &= V_\theta^l(\theta_l^*) - V_\theta^m(\theta_l^*); \\ \frac{\partial^2 \mathcal{L}}{\partial \theta_h^2} &= V_\theta^m(\theta_h^*) - V_\theta^h(\theta_h^*); \\ \frac{\partial^2 \mathcal{L}}{\partial \theta_l \partial \theta_h} &= 0; \\ \frac{\partial^2 \mathcal{L}}{\partial \theta_l \partial \lambda} &= f(\theta_l^*)(c_m - c_l); \\ \frac{\partial^2 \mathcal{L}}{\partial \theta_h \partial \lambda} &= f(\theta_h^*)(c_h - c_m). \end{aligned}$$

Then, $\det(M) > 0 \Leftrightarrow -(\frac{\partial^2 \mathcal{L}}{\partial \theta_l \partial \lambda})^2 \frac{\partial^2 \mathcal{L}}{\partial \theta_h^2} - (\frac{\partial^2 \mathcal{L}}{\partial \theta_h \partial \lambda})^2 \frac{\partial^2 \mathcal{L}}{\partial \theta_l^2} > 0$. Therefore, the sufficient conditions for the critical values that solve (P1) are,

$$f(\theta_l^*)^2 (c_m - c_l)^2 [V_\theta^h(\theta_h^*) - V_\theta^m(\theta_h^*)] > -f(\theta_h^*)^2 (c_h - c_m)^2 [V_\theta^m(\theta_l^*) - V_\theta^l(\theta_l^*)], \quad (\text{A4})$$

and will always be satisfied because under Assumption 2 $V_\theta^h(\theta_h) > V_\theta^m(\theta_h)$ and $V_\theta^m(\theta_l) > V_\theta^l(\theta_l)$. ■

Necessary conditions for interior solutions to (P1) The following conditions are necessary for an “interior solution” of (P1) in which all three treatments l, m and h (in ascending order of the distribution of θ) are offered:

- i) $V^l(\underline{\theta}) - c_l > V^m(\underline{\theta}) - c_m > V^h(\underline{\theta}) - c_h$;
- ii) $V^l(\bar{\theta}) - c_l < V^m(\bar{\theta}) - c_m < V^h(\bar{\theta}) - c_h$;
- iii) $V^h(\theta_l) - c_h < V^l(\theta_l^*) - c_l$.
- iv) $V^h(\theta_l) - c_h < V^m(\theta_l^*) - c_m$. ■

Changes in budget The following proposition indicates the effects of changes in available budget to the county, B , on P_N, P_L and P_H .

Proposition 1 *If the budget B increases then:*

- a) *the total proportion of people not trained, P_N , decreases;*
- b) *the proportion of people receiving training H , P_H , increases;*
- c) *the change in the proportion of people receiving training L , P_L , is ambiguous in Case 1 and Case 6, negative in Case 3 and Case 5, and positive in Case 2 and Case 4.*

Proof. Equations (A2) and the third line of (A1) form a system of implicit equations

$$\Pi_1(\theta_l, \theta_h, B, \Gamma, \Psi) = \frac{V^h(\theta_h; \Gamma) - V^m(\theta_h; \Gamma)}{c_h - c_m} - \frac{V^m(\theta_l; \Gamma) - V^l(\theta_l; \Gamma)}{c_m - c_l} = 0 \quad (\text{A5})$$

$$\Pi_2(\theta_l, \theta_h, B, \Gamma, \Psi) = B + F(\theta_l; \Psi)(c_m - c_l) + F(\theta_h; \Psi)(c_h - c_m) - c_h = 0. \quad (\text{A6})$$

Differentiating (A5) and (A6) with respect to θ_l, θ_h and B , and reordering terms, the following system needs to be solved to obtain the effect of changes of B on θ_l and θ_h :

$$\begin{bmatrix} d\theta_l \\ d\theta_h \end{bmatrix} = A^{-1} \begin{bmatrix} 0 \\ dB \end{bmatrix}, \quad (\text{A7})$$

$$\text{where } A = \begin{bmatrix} -\frac{V_\theta^m(\theta_l) - V_\theta^l(\theta_l)}{c_m - c_l} & \frac{V_\theta^h(\theta_h) - V_\theta^m(\theta_h)}{c_h - c_m} \\ -f(\theta_l)(c_m - c_l) & -f(\theta_h)(c_h - c_m) \end{bmatrix}. \quad (\text{A8})$$

$$\text{Note that } A^{-1} = \frac{1}{\det(A)} \begin{bmatrix} -f(\theta_h)(c_h - c_m) & -\frac{V_\theta^h(\theta_h) - V_\theta^m(\theta_h)}{c_h - c_m} \\ f(\theta_l)(c_m - c_l) & -\frac{V_\theta^m(\theta_l) - V_\theta^l(\theta_l)}{c_m - c_l} \end{bmatrix} \quad (\text{A9})$$

$$\text{where } \det(A) = \frac{V_\theta^m(\theta_l) - V_\theta^l(\theta_l)}{c_m - c_l} f(\theta_h)(c_h - c_m) + \frac{V_\theta^h(\theta_h) - V_\theta^m(\theta_h)}{c_h - c_m} f(\theta_l)(c_m - c_l). \quad (\text{A10})$$

The sign of this determinant is key to evaluate the effects of changes in B , and it will depend on the particular case under which the county is operating. Because of Assumption 2 it always holds that $V_\theta^m(\theta_l) - V_\theta^l(\theta_l) > 0$ and that

$V_\theta^h(\theta_h) - V_\theta^m(\theta_h) > 0$. Therefore, the sign of the determinant will depend on the differences of costs in each case. Given that $c_H > c_L > c_N$, it is easy to see that under Case 1 $c_h > c_m > c_l$; under Case 2 and Case 4 $c_m > c_l$ and $c_h < c_m$; under Case 3 and Case 5 $c_m < c_l$ and $c_h > c_m$; and under Case 6 $c_h < c_m < c_l$. In this way $\det(A)$ will be positive in Cases 1 and 6, and negative in Cases 2, 3, 4, and 5. ■

Hence, using (A7), (A9) and (A10), the effect of changes in B on the critical values θ_l and θ_h will be:

$$\frac{\partial \theta_l}{\partial B} = -\frac{1}{\det(A)} \frac{V_\theta^h(\theta_h) - V_\theta^m(\theta_h)}{c_h - c_m} \quad (\text{A11})$$

$$\frac{\partial \theta_h}{\partial B} = -\frac{1}{\det(A)} \frac{V_\theta^m(\theta_l) - V_\theta^l(\theta_l)}{c_m - c_l} \quad (\text{A12})$$

where, given the sign of the cost differences and of $\det(A)$ in each case, it is clear that $\frac{\partial \theta_l}{\partial B}$ will be negative in Cases 1, 2 and 4, and it will be positive in Cases 3, 5 and 6. Likewise $\frac{\partial \theta_h}{\partial B}$ will be negative in Cases 1, 3 and 5, and it will be positive in Cases 2, 4 and 6.

Finally, analyzing for each case P_l , P_m , and P_h , the proportion of individuals receiving treatment l , m , and h respectively, it is easy to see that:

$$\frac{\partial P_l}{\partial B} = f(\theta_l) \frac{\partial \theta_l}{\partial B} \quad (\text{A13})$$

$$\frac{\partial P_m}{\partial B} = f(\theta_h) \frac{\partial \theta_h}{\partial B} - f(\theta_l) \frac{\partial \theta_l}{\partial B} \quad (\text{A14})$$

$$\frac{\partial P_h}{\partial B} = -f(\theta_h) \frac{\partial \theta_h}{\partial B}. \quad (\text{A15})$$

Using (A11) and (A12) in (A13), (A14) and (A15), it can be seen then that the proportion of people receiving treatment N will always decrease (i.e. total training will increase) ($\frac{\partial P_N}{\partial B} < 0 \Leftrightarrow \frac{\partial(P_L+P_H)}{\partial B} > 0$), and that the proportion of people receiving training H will always increase ($\frac{\partial P_H}{\partial B} > 0$). The change in the proportion of people receiving training L will be negative in Cases 2 and 4 (where training H is offered to the individuals in the middle of the distribution of θ) and positive in Cases 3 and 5 (where treatment N is offered to the individuals in the middle of the distribution of θ). However, it will be ambiguous in Cases 1 and 6 (where treatment L is offered to the individuals in the middle of the distribution of θ), because in the equation $\frac{\partial P_L}{\partial B} = f(\theta_h) \frac{\partial \theta_h}{\partial B} - f(\theta_l) \frac{\partial \theta_l}{\partial B}$ (even without taking into account $f(\theta_l)$ and $f(\theta_h)$), it is ambiguous whether $\frac{\partial \theta_l}{\partial B} > \frac{\partial \theta_h}{\partial B}$ or $\frac{\partial \theta_l}{\partial B} < \frac{\partial \theta_h}{\partial B}$. ■

Effects of changes in the opportunity cost of training One can parameterize Y^N with a vector Γ that measures changes in the outcome generating function under no treatment in two ways: by increasing Y^N in a constant manner, and by increasing it more for more able individuals. Then, $Y^N(\theta; \Gamma) = \gamma_0 + \gamma_1 Y^N(\theta)$, where the vector Γ is formed by two elements: γ_0 which implies a parallel translation of the original function $Y^N(\theta)$ and γ_1 which implies a proportional increase of the returns to skills. Note that a change in Y^N affects also Y^L and Y^H (the outcome under training L and H), but just through the effect of Y^N . This rules out the possibility that changes in the economic situation of the county also affect the treatment effects Δ^L or Δ^H .

The effects of a change in γ_0 , a parallel translation in Y^N , are specified in the following proposition, and depend on a condition that is fully explained below.⁴⁷

Proposition 2 a) In general the effects of a parallel translation in the outcome generating function Y^N are ambiguous;

b) If the ratio of the opportunity cost to its direct cost of training H is “small enough” relative to the same ratio for training L , then an increase in γ_0 will reduce P_L and will increase P_N and P_H .

Proof. Differentiating (A5) and (A6) with respect to θ_l, θ_h and γ_0 , and reordering terms, the following system

$$\begin{bmatrix} d\theta_l \\ d\theta_h \end{bmatrix} = A^{-1} \begin{bmatrix} \left(\frac{\kappa_m - \kappa_l}{c_m - c_l} - \frac{\kappa_h - \kappa_m}{c_h - c_m} \right) d\gamma_0 \\ 0 \end{bmatrix}, \quad (\text{A16})$$

where A is defined as in (A8), needs to be solved to obtain the effects of changes of γ_0 on θ_l and θ_h . Note that $\kappa_m - \kappa_l \equiv -(\tau_m - \tau_l)$, $\kappa_h - \kappa_m \equiv -(\tau_h - \tau_m)$, for $l, m, h = N, L, H$, and $\tau_N \equiv 0$.⁴⁸ Then, replacing these terms

⁴⁷The effects of a change in γ_1 , a proportional translation in Y^N , are the same as the effects of a change in γ_0 . The proposition proving this, available upon request, is not presented in the interest of space.

⁴⁸In the case in which the discount factor is not one, the terms will be a little bit different, but all the analysis holds the same. Specifically, with $\rho < 1$, $\kappa_m - \kappa_l \equiv -\frac{\rho^{\tau_m} - \rho^{\tau_l}}{1 - \rho}$, $\kappa_h - \kappa_m \equiv -\frac{\rho^{\tau_h} - \rho^{\tau_l}}{1 - \rho}$.

and using (A9), (A10) and (A16), the effects can be expressed as

$$\frac{\partial \theta_l}{\partial \gamma_0} = -\frac{1}{\det(A)} f(\theta_h)(c_h - c_m) \left[\frac{\tau_h - \tau_m}{c_h - c_m} - \frac{\tau_m - \tau_l}{c_m - c_l} \right] \quad (A17)$$

$$\frac{\partial \theta_h}{\partial \gamma_0} = \frac{1}{\det(A)} f(\theta_l)(c_m - c_l) \left[\frac{\tau_h - \tau_m}{c_h - c_m} - \frac{\tau_m - \tau_l}{c_m - c_l} \right], \quad (A18)$$

where, by the analysis made in the proof of Proposition 1, $\det(A)$ will be positive in Cases 1 and 6 and negative in Cases 2 through 5. It is clear that the sign of the derivatives will depend crucially on the sign of the term between square brackets. Each ratio in the bracketed term represents the ratio of relative opportunity costs of two treatments (the higher τ , the longer the number of periods in which the individual can not participate in the labor market) versus the relative direct costs of these two treatments (as $c_H > c_L > c_N$ and $\tau_H > \tau_L > \tau_N$, the two ratios will be always positive). ■

To analyze the changes in the proportions, use expressions analogous to (A13), (A14) and (A15), and introduce (A17) and (A18) to get (rearranging terms):

$$\frac{\partial P_l}{\partial \gamma_0} = -\frac{1}{\det(A)} f(\theta_l) f(\theta_h)(c_h - c_m) \left[\frac{\tau_h - \tau_m}{c_h - c_m} - \frac{\tau_m - \tau_l}{c_m - c_l} \right] \quad (A19)$$

$$\frac{\partial P_m}{\partial \gamma_0} = \frac{1}{\det(A)} f(\theta_l) f(\theta_h)(c_h - c_l) \left[\frac{\tau_h - \tau_m}{c_h - c_m} - \frac{\tau_m - \tau_l}{c_m - c_l} \right] \quad (A20)$$

$$\frac{\partial P_h}{\partial \gamma_0} = -\frac{1}{\det(A)} f(\theta_l) f(\theta_h)(c_m - c_l) \left[\frac{\tau_h - \tau_m}{c_h - c_m} - \frac{\tau_m - \tau_l}{c_m - c_l} \right]. \quad (A21)$$

From (A19), (A20) and (A21), it is clear that each proportion change according to a weight given by the difference in direct costs of the other two treatments, times the difference in ratios of opportunity costs. The sign of these differences will be different in each case, but it is straightforward to see that if

$$\frac{\tau_H}{c_H} < \frac{\tau_L}{c_L},$$

the bracketed expression will be negative in Cases 1, 2 and 3, and it will be positive in Cases 4, 5 and 6. This implies that in every case $\frac{\partial P_N}{\partial \gamma_0} > 0$, $\frac{\partial P_L}{\partial \gamma_0} < 0$ and $\frac{\partial P_H}{\partial \gamma_0} > 0$.⁴⁹ ■

This proposition shows that a change in the opportunity cost of training will have two opposite effects. On one hand, there is a direct effect of increasing the cost of no treatment, which implies that P_N should always increase. How P_L and P_H change is more complex to analyze. Given that the budget is fixed and that P_N increases, with everything else equal, the county has now more budget available for the people that are going to be trained. Following the same logic of Proposition 1, given that training H is always rationed, there should be a substitution of training L for training H (therefore making P_L decrease and P_H increase). However, because training H is the training that takes longer to be completed, the opportunity cost of this training with respect to training N increases more than for training L , making training H less desirable. If this effect is big enough, then the substitution will be from training H to training L , which in turn might even imply (if the savings are big enough, given that training H is more expensive) that the increase in P_L makes P_N actually decrease.

Effects of change in skills distribution Let's call Ψ the vector of parameters that define the distribution of skills, and represent this distribution as $F(\theta; \Psi)$. Using the definitions of the proportions of individuals receiving treatment l, m and h , it is easy to see that the effect on these proportion of a change in Ψ will be composed by two parts: the change in the optimal θ_l and θ_h because of the distributional change, and the change in the mass of the distribution for given θ_l and θ_h :

$$\frac{\partial P_l}{\partial \Psi} = f(\theta_l) \frac{\partial \theta_l}{\partial \Psi} + F_\Psi(\theta_l) \quad (A22)$$

$$\frac{\partial P_m}{\partial \Psi} = [f(\theta_h) \frac{\partial \theta_h}{\partial \Psi} + F_\Psi(\theta_h)] - [f(\theta_l) \frac{\partial \theta_l}{\partial \Psi} + F_\Psi(\theta_l)] \quad (A23)$$

$$\frac{\partial P_h}{\partial \Psi} = -[f(\theta_h) \frac{\partial \theta_h}{\partial \Psi} + F_\Psi(\theta_h)]. \quad (A24)$$

The effects of a change in Ψ can be analyzed in the same way as with changes in B or γ_0 . In particular,

⁴⁹With a discount factor $\rho < 1$, the analogous condition (which has the same economic implications) will be $\frac{1-\rho^{\tau_2}}{c_2} < \frac{1-\rho^{\tau_1}}{c_1}$.

differentiating (A5) and (A6) with respect to θ_l, θ_h and Ψ , and reordering terms, the following system

$$\begin{bmatrix} d\theta_l \\ d\theta_h \end{bmatrix} = A^{-1} \begin{bmatrix} 0 \\ [(c_m - c_l)F_\Psi(\theta_l) + (c_h - c_m)F_\Psi(\theta_h)]d\Psi \end{bmatrix}, \quad (\text{A25})$$

where A is defined as in (A8), needs to be solved to obtain the effects of changes of Ψ on θ_l and θ_h . Then, using (A9), (A10), (A11), (A12) and (A25), the effect of changes in Ψ on the critical values θ_l and θ_h can be expressed as:

$$\frac{\partial\theta_l}{\partial\Psi} = \frac{\partial\theta_l}{\partial B} [(c_m - c_l)F_\Psi(\theta_l) + (c_h - c_m)F_\Psi(\theta_h)] \quad (\text{A26})$$

$$\frac{\partial\theta_h}{\partial\Psi} = \frac{\partial\theta_h}{\partial B} [(c_m - c_l)F_\Psi(\theta_l) + (c_h - c_m)F_\Psi(\theta_h)]. \quad (\text{A27})$$

The term in the square bracket shows that the interaction of the relative costs of the treatments and the changes in the mass of the distribution of θ , will determine the effect on θ_l and θ_h . However, (A26) and (A27) are not enough to characterize the effects of changes in Ψ on the proportions, because as it is shown in (A22), (A23) and (A24), there will be an extra term, F_Ψ that will make the final effect in most of the cases ambiguous (many of the signs would be ambiguous even without this extra term).

I make two different distributional assumptions. The first is assume that $F(\theta)$ is distributed Uniform $(\theta, \bar{\theta})$, and define $\Psi \equiv \bar{\theta}$. This implies studying the effect of the entrance (or exit) of more able individuals (given the properties of the Uniform distribution, an increase in $\bar{\theta}$ imply an increase in both $E(\theta)$ and $V(\theta)$). The results are ambiguous in most cases: $F_{\bar{\theta}}$ is always negative, but $\frac{\partial\theta_l}{\partial\bar{\theta}}$ and $\frac{\partial\theta_h}{\partial\bar{\theta}}$ are some times positive and some times negative (depending on the case). This reflects the fact that depending on the (relative) cost of the treatment towards which the distribution moves, the county would be able to increase the proportion of people receiving that particular treatment. The results below show that in the cases (except Case 1, ambiguous) in which training L is offered to individuals of relative less skills than the ones offered training H , P_L decreases unambiguously if $\bar{\theta}$ increases. Intuitively, given that the budget remains constant, one should think that this would imply a less than proportional increase in P_H and an increase in P_N . In the same way, intuition suggests that the opposite effects should occur for cases in which training L is offered to (relatively) more able individuals than individuals offered training H .

The second assumption is that $F(\theta)$ is distributed Normal (μ, σ^2) , and $\Psi \equiv [\mu, \sigma^2]$. Below I show that analyzing the effects of changes in μ and in σ^2 gives also ambiguous results. However, because an increase in μ with σ^2 fixed implies that the new distribution stochastically dominates the old one, the intuition would be equivalent to changing $\bar{\theta}$ in the Uniform case.

Uniform case Assume θ is distributed Uniform $(\theta, \bar{\theta})$, and define $\Psi = \bar{\theta}$, then $F(\theta) = \frac{\theta - \theta}{\bar{\theta} - \theta}$, $f(\theta) = \frac{1}{\bar{\theta} - \theta}$, which implies that $F_{\bar{\theta}}(\theta) = -\frac{\theta - \theta}{(\bar{\theta} - \theta)^2} = -f(\theta)F(\theta) < 0$ and that $\frac{\partial F_{\bar{\theta}}(\theta)}{\partial\bar{\theta}} = -f(\theta)^2 < 0$ (that is, an increase in $\bar{\theta}$ implies that the new distribution stochastically dominates the old one, and that the distance between the two cumulative distribution functions increases with θ). The bracketed term in (A23) and (A24) will be negative in Cases 1 and 3, positive in Cases 4 and 6, and will have an ambiguous sign in Cases 2 and 5, which implies that in Cases 1 and 6 $\frac{\partial\theta_l}{\partial\bar{\theta}} > 0$ and $\frac{\partial\theta_h}{\partial\bar{\theta}} > 0$, in Cases 3 and 4 $\frac{\partial\theta_l}{\partial\bar{\theta}} < 0$ and $\frac{\partial\theta_h}{\partial\bar{\theta}} > 0$ and in Cases 2 and 5 the derivatives will have an ambiguous sign. Using (A22), (A23) and (A24), the only unambiguous signs will be $\frac{\partial P_L}{\partial\bar{\theta}} < 0$ in Cases 3 and 4.

Normal case Assume θ is distributed Normal (μ, σ^2) , and define $\Psi = [\mu, \sigma^2]$. Then for a particular value of θ , say $\tilde{\theta}$, $F_\mu(\tilde{\theta}) = \int_{-\infty}^{\tilde{\theta}} f(\theta) \left(\frac{\theta - \mu}{\sigma^2}\right) d\theta = \int_{-\infty}^{\frac{\tilde{\theta} - \mu}{\sigma}} \phi\left(\frac{\tilde{\theta} - \mu}{\sigma}\right) z dz < 0$, where ϕ is the pdf of a standard Normal, and $z \equiv \frac{\theta - \mu}{\sigma}$. Note that although $F_\mu(\tilde{\theta})$ is negative, it will attain a minimum at $\tilde{\theta} = \mu$ and then it will increase asymptotically towards zero with θ . Therefore, when evaluating $F_\mu(\theta_l)$ and $F_\mu(\theta_h)$ in (A22) and (A23), it is not possible to determine which value is higher. However, still an increase in μ implies that the new distribution stochastically dominates the old one, and intuitively results are similar to changes in $\bar{\theta}$ in the Uniform case.

Changes in σ^2 are even more difficult to evaluate because $F_{\sigma^2}(\tilde{\theta}) = \frac{1}{2\sigma^2} \int_{-\infty}^{\tilde{\theta}} f(\tilde{\theta}) \left(\left(\frac{\theta - \mu}{\sigma}\right)^2 - 1\right) d\theta = \frac{1}{2\sigma^2} \left[\int_{-\infty}^{\frac{\tilde{\theta} - \mu}{\sigma}} \phi\left(\frac{\tilde{\theta} - \mu}{\sigma}\right) z^2 dz - F(\tilde{\theta})\right] \lesseqgtr 0$ (it is > 0 when $\tilde{\theta} < \mu$, $= 0$ when $\tilde{\theta} = \mu$, and < 0 when $\tilde{\theta} > \mu$). Hence, it is not possible to evaluate $F_{\sigma^2}(\theta_l)$ versus $F_{\sigma^2}(\theta_h)$ in (A22) and (A23), given that the position of θ_l and θ_h with respect to μ is unknown. ■

Preference for one type of training The solution to (P2) is similar to the solution to (P1). The first order conditions from (P2) are:

$$\theta_l : -\alpha[V^L(\theta_l) - V^N(\theta_l) - c_L]f(\theta_l) + \lambda f(\theta_l)c_L \leq 0 \quad (\text{A28})$$

$$\theta_h : \{\alpha[V^L(\theta_h) - c_L] - (1 - \alpha)[V^H(\theta_h) - c_H] + (1 - 2\alpha)V^N(\theta_h)\}f(\theta_h) + \lambda f(\theta_h)(c_H - c_L) \leq 0 \quad (\text{A29})$$

$$\lambda : B - [F(\theta_h) - F(\theta_l)]c_L - [1 - F(\theta_h)]c_H \geq 0. \quad (\text{A30})$$

Combining (A28) and (A29), and defining $a \equiv \frac{1-\alpha}{\alpha}$ as the preference factor for training H , the agency chooses θ_l and θ_h such that the following condition is satisfied:

$$\frac{V^L(\theta_l) - V^N(\theta_l)}{c_L} = \frac{aV^H(\theta_h) - V^L(\theta_h) - (a-1)[V^N(\theta_h) - c_H]}{c_H - c_L} \quad (\text{A31})$$

That is, as with (P1) the ratio of marginal benefits to marginal costs of each training are equated, but with a weight factor to benefits given by a . It is easy to see that when $\alpha = 1/2$ (i.e. there is equal weight given to each type of training), then $a = 1$, and (A31) collapses to the optimality condition (A2).

If α is such that an interior solution occurs in which (A31) is satisfied, then the empirical implications regarding budget and opportunity cost of training under (P1) still hold, and empirical implications can be derived for changes in α . For that is necessary that the preference for training L α be in a certain range, as is specified in the following proposition:

Proposition 3 *If the preference for training L is such that $\alpha < \frac{V^H(\theta_h) - V^N(\theta_h)}{V^H(\theta_h) + V^L(\theta_h) - 2V^N(\theta_h)}$ then:*

a) *the results in Proposition 1 hold;*

b) *if also $\alpha > \frac{\kappa_N - \kappa_H}{\frac{c_H - c_L}{c_L} \kappa_N - \frac{c_H}{c_L} \kappa_L + \kappa_H}$ then the results of Proposition 2 hold;*

c) *an increase in α increases P_L and decreases P_N and P_H .*

Proof. The condition $\alpha < \frac{V^H(\theta_h) - V^N(\theta_h)}{V^H(\theta_h) + V^L(\theta_h) - 2V^N(\theta_h)}$ is a sufficient condition for the second order conditions of (P2) to be satisfied (because it assures that $[aV_\theta^H(\theta_h) - V_\theta^L(\theta_h) - (a-1)V_\theta^N(\theta_h)] > 0$, which is a condition analogous to the one imposed in Assumption 2). Then, (A30) and (A31) can be used to form the following system of equations:

$$\Pi_1(\theta_l, \theta_h, B, \gamma_0) = \frac{V^L(\theta_l) - V^N(\theta_l)}{c_L} - \frac{aV^H(\theta_h) - V^L(\theta_h) - (a-1)[V^N(\theta_h) - c_H]}{c_H - c_L} = 0 \quad (\text{A32})$$

$$\Pi_2(\theta_l, \theta_h, B, \gamma_0) = B - [F(\theta_h) - F(\theta_l)]c_L - [1 - F(\theta_h)]c_H = 0. \quad (\text{A33})$$

Part a) can be proved by differentiating (A32) and (A33) with respect to θ_l, θ_h and B , and reordering terms to form the following system

$$\begin{bmatrix} d\theta_l \\ d\theta_h \end{bmatrix} = A^{-1} \begin{bmatrix} 0 \\ dB \end{bmatrix}, \quad (\text{A34})$$

where

$$A^{-1} = \frac{1}{\det(A)} \begin{bmatrix} -f(\theta_h)(c_H - c_L) & -[aV_\theta^H(\theta_h) - V_\theta^L(\theta_h) - (a-1)V_\theta^N(\theta_h)] \\ f(\theta_l)c_L & -\frac{V_\theta^L(\theta_l) - V_\theta^N(\theta_l)}{c_L} \end{bmatrix} \quad (\text{A35})$$

and

$$\det(A) = \frac{V_\theta^L(\theta_l) - V_\theta^N(\theta_l)}{c_L} f(\theta_h)(c_H - c_L) + [aV_\theta^H(\theta_h) - V_\theta^L(\theta_h) - (a-1)V_\theta^N(\theta_h)]f(\theta_l)c_L. \quad (\text{A36})$$

The sign of the determinant in (A36) is positive, and it is straightforward to see that the same results with respect to changes in B as under (P1) hold (see (A11) through (A15)).

Part b) can be proved by differentiating (A32) and (A33) with respect to θ_l, θ_h and γ_0 , and using (A35) and (A36) in the system

$$\begin{bmatrix} d\theta_l \\ d\theta_h \end{bmatrix} = A^{-1} \begin{bmatrix} \left(\frac{\kappa_L - \kappa_N}{c_L} - \frac{a\kappa_H - \kappa_L - (a-1)\kappa_N}{c_H - c_L} \right) d\gamma_0 \\ 0 \end{bmatrix} \quad (\text{A37})$$

where under the condition imposed in part b) it can be shown that $\frac{\kappa_L - \kappa_N}{c_L} < \frac{a\kappa_H - \kappa_L - (a-1)\kappa_N}{c_H - c_L}$ and all the results in (A17) through (A21) hold.

Finally, part c) of the proposition can be proved by differentiating (A32) and (A33) with respect to θ_l, θ_h and a , and forming the system

$$\begin{bmatrix} d\theta_l \\ d\theta_h \end{bmatrix} = A^{-1} \begin{bmatrix} \frac{V^H(\theta_h) - V^N(\theta_h) - c_H}{c_H - c_L} da \\ 0 \end{bmatrix} \quad (\text{A38})$$

where (A35) and (A36) still hold. Solving (A38), it is clear that

$$\frac{\partial \theta_l}{\partial a} = \frac{1}{\det(A)} f(\theta_h) (c_H - c_L) \frac{V^H(\theta_h) - V^N(\theta_h) - c_H}{c_H - c_L} \quad (\text{A39})$$

$$\frac{\partial \theta_h}{\partial a} = -\frac{1}{\det(A)} f(\theta_l) c_L \frac{V^H(\theta_h) - V^N(\theta_h) - c_H}{c_H - c_L} \quad (\text{A40})$$

will be positive and negative respectively. This clearly implies that as a increases P_L decreases and P_N and P_H increase (see for example (A13)-(A15)). Given that $a = \frac{1-\alpha}{\alpha}$, then α and a will go in opposite directions, and an increase in α will increase P_L and decrease P_N and P_H . ■

Objective function with inequality aversion The problem in which the program administrator exhibits inequality aversion can be expressed using a Constant Elasticity of Substitution (CES) welfare function for a decision maker that cares about the after treatment outcome distribution and that solves the following problem:

$$\begin{aligned} \max_{\{\theta_l, \theta_h\}} W &= \frac{1}{\varepsilon} \left[\int_{\underline{\theta}}^{\theta_l} [V^l(\theta)]^\varepsilon dF(\theta) + \int_{\theta_l}^{\theta_h} [V^m(\theta)]^\varepsilon dF(\theta) + \int_{\theta_h}^{\bar{\theta}} [V^h(\theta)]^\varepsilon dF(\theta) \right] \\ \text{s.t. } &F(\theta_l) c_l + [F(\theta_h) - F(\theta_l)] c_m + [1 - F(\theta_h)] c_h \leq B, \end{aligned} \quad (\text{P3})$$

where ε is the inequality aversion parameter. When $\varepsilon = 1$ there is no inequality aversion, if $\varepsilon = 0$ the utility function becomes logarithmic and implies unitary inequality aversion, and if $\varepsilon \rightarrow -\infty$ then there infinite aversion to inequality and the welfare function becomes Rawlsian where only the welfare of the individual worst off matters. Solving for the first order conditions of this problem it is straightforward to show that the optimality condition will be

$$\frac{[V^h(\theta_h)]^\varepsilon - [V^m(\theta_h)]^\varepsilon}{[V^m(\theta_l)]^\varepsilon - [V^l(\theta_l)]^\varepsilon} = \frac{c_h - c_m}{c_m - c_l} \quad (\text{A41})$$

which is analogous to the first order condition of (P1), equation (A2). In fact (P1) is a special case of (P3) when $\varepsilon = 1$, i.e. there is no aversion to inequality. It can be shown that for some low levels of inequality aversion the results of (P1) will hold, but for higher levels of inequality aversion the results can actually be very different. If the first case occurs, though, it would not be possible to differentiate between preferences towards the lower earnings group introduced by the inequality aversion parameter, from the preferences for a type of training (as in (P2)). If the inequality aversion is high, then depending which training is more beneficial for the lower skills individuals, the results from (P1) could be even stronger, or could become completely the opposite. In particular, if training H favors lower skills individuals, the results from (P3) will be similar with respect to changes in budget and local economic conditions to the results from (P1). ■

References

- [1] Bean, Clive and Elim Papadakis (1998). "A Comparison of Mass Attitudes Towards the Welfare State in Different Institutional Regimes, 1985-1990." *International Journal of Public Opinion Research*, Volume 10, Issue 3, (Fall), 211-227.
- [2] Behncke, Stefanie, Markus Frölich, and Michael Lechner (2007). "Targeting Labour Market Programmes - Results from a Randomized Experiment", IZA Discussion Paper No. 3085.
- [3] Bell, Stephen and Larry Orr (2002). "Screening (and creaming?) applicants to job training programs: the AFDC homemaker-home health aide demonstrations", *Labour Economics* Volume 9, Issue 2 (April), 279-301.
- [4] Berger, Mark, Dan Black, and Jeffrey Smith (2001). "Evaluating Profiling as a Means of Allocating Government Services" in *Econometric Evaluation of Labour Market Policies*, Lechner, Michael and Friedhelm Pfeiffer (eds.), Physica-Verlag Heidelberg.
- [5] Bloom, Dan, and Charles Michalopoulos (2001). How Welfare and Work Policies Affect Employment and Income, MDRC, <http://www.mdrc.org/publications/99/overview.html> (Accessed May 17, 2008).
- [6] Bloom, Howard S., Carolyn J. Hill, and James A. Riccio (2003). "Linking Program Implementation and Effectiveness: Lessons from a Pooled Sample of Welfare-to-Work Experiments", *Journal of Policy Analysis and Management*, Volume 22, No. 4 (Fall), 551-575.
- [7] Blough, David K., Carolyn W. Madden, and Mark C. Hornbrook (1999). "Modeling risk using generalized linear models", *Journal of Health Economics*, Volume 18, No. 2, 153-171.
- [8] CDSS (2005). "TANF High Performance Bonus: California's Accomplishments", California Department of Social Services, November, http://www.dss.cahwnet.gov/research/res/pdf/HighPerfBonus10_05.pdf (accessed December 12, 2008).
- [9] Cook, Fay and Edith Barrett (1992). *Support for the American Welfare State: the Views of Congress and the Public*, Columbia University Press.
- [10] Courty, Pascal and Gerald Marschke (1997). "Measuring Government Performance: Lessons from a Federal Job-Training Program", *American Economic Review*, Volume 87, Issue 2, 383-388.
- [11] Courty, Pascal, and Gerald Marschke (2004). "An Empirical Investigation of Gaming Responses to Explicit Performance Incentives", *Journal of Labor Economics*, Volume 22, No. 1, 23-56.
- [12] Courty, Pascal, Carolyn Heinrich, and Gerald Marschke (2005). "Setting the Standard in Performance Measurement Systems", *International Public Management Journal*, Volume 8, No. 3, 321-347.
- [13] Dehejia, Rajeev (2005). "Program Evaluation as a Decision Problem", *Journal of Econometrics*, Volume 125, Issues 1-2 (March-April), 141-173.
- [14] Dixit, Avinash (2002). "Incentives and Organization in the Public Sector", *Journal of Human Resources*, Volume 87, No. 4 (Fall), 697-727.

- [15] Dyke, Andrew, Carolyn J. Heinrich, Peter R. Mueser, and Kyung-Seong Jeon (2006). "The Effects of Welfare-to-Work Program Activities on Labor Market Outcomes", *Journal of Labor Economics*, Volume 24, No. 3, 567-607.
- [16] Eberts, Randall, Kevin Hollenbeck, and Joe Stone (2002). "Teacher Performance Incentives and Student Outcomes", *Journal of Human Resources*, Volume 87, No. 4 (Fall), 913-927.
- [17] Erikson, Robert S., Michael B. MacKuen, and James A. Stimson (2002). "Panderers or Shirkers: Politicians and Public Policy", in *Navigating Public Opinion: Polls, Policy, and the Future of American Democracy*, J. Manza, F. Cook and B. Page, ed., Oxford University Press.
- [18] Fording, Richard, Sanford Schram and Joe Soss (2006). "The Bottom Line, the Business Model and the Bogy: Performance Management, Sanctions and the Brave New World of Welfare-to-Work in Florida", UKCPR Discussion Paper DP 2006-10 (December).
- [19] Fording, Richard, Joe Soss and Sanford Schram (2007). "Devolution, Discretion and the Effect of Local Political Values on TANF Sanctioning", *Social Service Review*, Volume 81, No. 2 (June), 285-316.
- [20] Fording, Richard, Joe Soss and Sanford Schram (2008). "Distributing Discipline: Race, politics, and punishment at the frontlines of welfare reform", UKCPR Discussion Paper DP 2008-06 (November).
- [21] Grogger, Jeffrey, and Charles Michalopoulos (2003). "Welfare Dynamics under Time Limits", *Journal of Political Economy*, Volume 111, No. 3 (June 1), 530-554.
- [22] Hamilton, Gayle, Stephen Freedman, Lisa Gennetian, Charles Michalopoulos, Johanna Walter, Diana Adamas-Ciardullo, Anna Gassman-Pines, Sharon McGroder, Martha Zaslow, Jennifer Brooks, Surjeet Ahluwalia, with Electra Small and Bryan Ricchetti (2001). "How Effective are Different Welfare-to-Work Approaches? Five Year Adult and Child Impacts for Eleven Programs", National Evaluation of Welfare-to-Work Strategies, MDRC, December.
- [23] 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, 95-114.
- [24] Heckman, James, Carolyn Heinrich, and Jeffrey Smith (1997). "Assessing the Performance of Performance Standards in Public Bureaucracies", *American Economic Review*, Volume 87, Issue 2, 389-395.
- [25] Heckman, James, Carolyn Heinrich, and Jeffrey Smith (2002). "The Performance of Performance Standards", *Journal of Human Resources*, Volume 87, No. 4 (Fall), 778-811.
- [26] Heckman, James and Bo Honore (1990). "The Empirical Content of the Roy Model", *Econometrica*, Vol. 58, No. 5, 1121-49
- [27] Heckman, James, Jeffrey Smith and Christopher Taber (1996). "What Do Bureaucrats Do? The Effects of Performance Standards and Bureaucratic Preferences on Acceptance into the JTPA Program", in *Advances in the Study of Entrepreneurship, Innovation and Economic Growth: Reinventing Government and the Problem of Bureaucracy*, Volume 7, Libecap, Gary, ed., Greenwich, CT, JAI Press, 191-217.
- [28] Hill, Carolyn J. (2006). "Casework Job Design and Client Outcomes in Welfare-to-Work Offices", *Journal of Public Administration Research and Theory*, Volume 16, No.2, 263-288.

- [29] Hotz, V. Joseph, Guido Imbens, and Jacob Klerman (2006). “Evaluating the Differential Effects of Alternative Welfare-to-Work Training Components: A Reanalysis of the California GAIN Program”, *Journal of Labor Economics*, Volume 24, No. 3, 521-566.
- [30] Hoynes, Hillary (2000). “Local Labor Markets and Welfare Spells: Do Demand Conditions Matter?”, *Review of Economics and Statistics*, Volume 82, No. 3 (August), 351-368.
- [31] Hoxby, Caroline (2002). “Would School Choice Change the Teaching Profession?”, *Journal of Human Resources*, Volume 87, No. 4 (Fall), 846-891.
- [32] Jacobs, Lawrence and Robert Shapiro (2002). “Politics and Policymaking in the Real World: Crafted Talk and the Loss of Democratic Responsiveness”, in *Navigating Public Opinion: Polls, Policy, and the Future of American Democracy*, J. Manza, F. Cook and B. Page, ed., Oxford University Press.
- [33] Keiser, Lael R., Peter R. Mueser, and Seung-Whan Choi (2004). “Race, Bureaucratic Discretion, and the Implementation of Welfare Reform”, *American Journal of Political Science*, Volume 48, No. 2 (April), 314-327.
- [34] Kluegel, James and Eliot Smith (1986). *Beliefs about Inequality*, Aldine de Gruyter, New York.
- [35] Koretz, Daniel (2002). “Limitations in the Use of Achievements Tests as Measures of Educators’ Productivity”, *Journal of Human Resources*, Volume 87, No. 4 (Fall), 752-778.
- [36] Lechner, Michael, and Jeffrey Smith (2007). “What is the value added by caseworkers?”, *Labour Economics*, Volume 14, No. 2, 135-151.
- [37] LAO (1996). “A Perspective on County Fiscal Constraints”, LAO Analysis of the 1996-97 Budget Bill, P&I Part 5-2, http://www.lao.ca.gov/analysis_1996/p965-2.html (accessed May 18, 2008).
- [38] LAO (2000). “The 2000-01 Budget: Perspectives and Issues”, Legislative Analyst’s Office Report, http://www.lao.ca.gov/analysis_2000/2000_pandi/pandi_2000.pdf (accessed December 12, 2008).
- [39] LAO (2001). “Improving CalWORKs Program Effectiveness: Changing the Employment Services Budget Process”, Legislative Analyst’s Office Report, http://www.lao.ca.gov/2001/calworks/021401_calworks.pdf (accessed May 18, 2008).
- [40] LAO (2002). “The 2002-03 Budget Bill: Perspectives and Issues”, Legislative Analyst’s Office Report, http://www.lao.ca.gov/analysis_2002/2002_pandi/pi_part_5c_calworks_anl02.html (accessed December 12, 2008).
- [41] Leung, Siu Fai and Shihti Yu (1996). “On the choice between sample selection and two-part models”, *Journal of Econometrics*, Volume 72, Issues 1-2 (May-June), 197-229.
- [42] Manning, Willard G., Naihua Duan, and William H. Rogers (1987). “Monte Carlo evidence on the choice between sample selection and two-part models”, *Journal of Econometrics*, Volume 35, No. 1, 59-82.
- [43] Manning, Willard G., and John Mullahy (2001). “Estimating log models: to transform or not to transform?”, *Journal of Health Economics*, Volume 20, No. 4, 461-494.

- [44] Manski, Charles (2000). "Identification problems and decisions under ambiguity: Empirical analysis of treatment response and normative analysis of treatment choice", *Journal of Econometrics*, Volume 95, Issue 2 (April), 415-442.
- [45] Manski, Charles (2004). "Statistical Treatment Rules for Heterogeneous Populations", *Econometrica*, Volume 72, Issue 4, 1221-1246.
- [46] Mullahy, John (1998). "Much ado about two: reconsidering retransformation and the two-part model in health econometrics", *Journal of Health Economics*, Volume 17, No. 3, 247-281.
- [47] Nathan, Richard and Thomas Gais (1998). "Implementing the Personal Responsibility Act of 1996: A First Look", Federalism Research Group, The Nelson A. Rockefeller Institute of Government, State University of New York.
- [48] Pepper, John (2002). "To Train or Not To Train: Optimal Treatment Assignment Rules Using Welfare-to-Work Experiments", University of Virginia (February).
- [49] Pepper, John (2003). "Using Experiments to Evaluate Performance Standards: What Do Welfare-to-Work Demonstration Reveal to Welfare Reformers?", *Journal of Human Resources*, Volume 38, No. 4, 860-80.
- [50] Ricucci, Norma (2005). "Street-Level Bureaucrats and Intrastate Variation in the Implementation of Temporary Assistance for Needy Families Policies", *Journal of Public Administration Research and Theory*, Volume 15, No. 1, 89-111.
- [51] Roy, A (1951). "Some Thoughts on the Distribution of Earnings". *Oxford Economic Papers*, Vol. 3, 135-146.
- [52] Salanié, Bernard (1997). *The economics of contracts: primer*. MIT Press.
- [53] Weaver, R. Kent (2002). "Polls, Priming, and the Politics of Welfare Reform", in *Navigating Public Opinion: Polls, Policy, and the Future of American Democracy*, J. Manza, F. Cook and B. Page, ed., Oxford University Press.
- [54] Willis, R. (1986). "Wage Determinants: A Survey and Reinterpretation of Human Capital Earnings Functions", in *Handbook of Labor Economics*, Volume I, O.Ashenfelter and R. Layard, ed., North-Holland.
- [55] Wiseman, Michael (2004). "The High Performance Bonus", George Washington Institute of Public Policy Working Paper No. 7, George Washington University (October).

Table 1. Evolution of county level variables (annual average 25 counties)

Variable	1994	1995	1996	1997	1998	1999 /a
Proportion Not Trained	0.796 (0.11)	0.780 (0.13)	0.753 (0.15)	0.685 (0.19)	0.484 (0.26)	0.287 (0.22)
Proportion Labor Force Attachment (LFA) Training	0.110 (0.07)	0.143 (0.08)	0.186 (0.11)	0.249 (0.14)	0.406 (0.22)	0.531 (0.21)
Proportion Human Capital Development (HCD) Training	0.093 (0.05)	0.077 (0.06)	0.061 (0.05)	0.066 (0.08)	0.110 (0.11)	0.182 (0.12)
Proportion Adult Basic Education (ABE) Training	0.064 (0.04)	0.054 (0.05)	0.039 (0.04)	0.037 (0.06)	0.038 (0.06)	0.055 (0.06)
Proportion Vocational (VOC) Training	0.028 (0.01)	0.024 (0.01)	0.021 (0.02)	0.029 (0.02)	0.071 (0.08)	0.128 (0.08)
Total Budget (\$ millions) /b	35.11 (33.33)	34.96 (32.71)	35.21 (33.11)	32.90 (30.57)	34.59 (30.71)	38.83 (33.40)
Training Budget (\$ millions) /b	5.17 (4.40)	5.18 (4.31)	5.35 (4.56)	6.05 (5.04)	8.70 (7.52)	11.03 (8.77)
Training Share of Total Budget	0.174 (0.05)	0.175 (0.05)	0.178 (0.05)	0.212 (0.06)	0.273 (0.06)	0.307 (0.05)
Proportion Democratic County Supervisors	0.479 (0.28)	0.451 (0.31)	0.462 (0.28)	0.450 (0.23)	0.451 (0.23)	0.443 (0.25)
Proportion Democratic Majority Board of Supervisors	0.561 (0.50)	0.602 (0.49)	0.602 (0.49)	0.577 (0.50)	0.579 (0.50)	0.554 (0.50)
Proportion >= 80% Democratic Supervisors ('strong')	0.125 (0.33)	0.113 (0.32)	0.113 (0.32)	0.060 (0.24)	0.060 (0.24)	0.076 (0.27)
Proportion = 60% Democratic Supervisors ('weak')	0.436 (0.50)	0.488 (0.50)	0.488 (0.50)	0.517 (0.50)	0.519 (0.50)	0.478 (0.50)
Unemployment Rate (%)	9.506 (3.46)	8.732 (3.77)	8.220 (3.79)	7.274 (3.75)	6.919 (3.87)	6.493 (4.01)
Employment Service Sector/Population	0.166 (0.05)	0.173 (0.05)	0.181 (0.05)	0.243 (0.05)	0.246 (0.06)	0.252 (0.06)
Average Real Earnings Service Sector/\$1,000	8.209 (1.65)	8.355 (1.92)	8.273 (1.82)	8.521 (1.63)	8.759 (1.63)	8.645 (1.54)
Enrollment Adults in Welfare (AFDC/TANF)	783,671	784,400	751,418	653,554	539,968	462,295

Notes:

Standard deviations between parentheses. Averages and standard deviations are weighted by average county welfare caseload in the period.

/a First two quarters only

/b Budget refers to data on total expenses (not including cash grants) and expenses in training in the welfare program (AFDC/TANF), where training refers to both GAIN (1994-1997) and WTW (1998-1999) programs.

Table 2. Characteristics new entrants (annual average 25 counties)

Variable	1994	1995	1996	1997	1998	1999 /a
Percentage Female Entrants	0.758 (0.06)	0.758 (0.06)	0.747 (0.06)	0.758 (0.06)	0.749 (0.06)	0.757 (0.05)
Average Age at Entry	27.7 (0.94)	27.0 (0.85)	26.6 (0.84)	26.2 (0.83)	25.8 (0.95)	25.6 (0.92)
Proportion White	0.339 (0.16)	0.350 (0.15)	0.350 (0.13)	0.355 (0.14)	0.335 (0.14)	0.327 (0.14)
Proportion Hispanic	0.421 (0.17)	0.399 (0.16)	0.394 (0.15)	0.378 (0.15)	0.373 (0.15)	0.371 (0.14)
Proportion Black	0.150 (0.08)	0.164 (0.09)	0.173 (0.09)	0.193 (0.10)	0.204 (0.10)	0.211 (0.10)
Proportion Other Races	0.089 (0.05)	0.087 (0.05)	0.083 (0.05)	0.075 (0.04)	0.088 (0.05)	0.090 (0.05)
Average Number of Kids at Entry	1.471 (0.12)	1.390 (0.08)	1.379 (0.09)	1.355 (0.08)	1.331 (0.09)	1.284 (0.09)
Proportion Entrants with Infants at Entry /b	0.216 (0.03)	0.232 (0.03)	0.245 (0.03)	0.267 (0.03)	0.287 (0.04)	0.288 (0.04)
Proportion Entrants with Toddlers at Entry /c	0.492 (0.03)	0.477 (0.02)	0.491 (0.04)	0.494 (0.04)	0.488 (0.03)	0.473 (0.04)
Average Quarters Employed 1 Yr before Entry	1.339 (0.17)	1.387 (0.15)	1.415 (0.16)	1.394 (0.14)	1.386 (0.13)	1.458 (0.13)
Average Quarters Employed 2 Yrs before Entry	1.085 (0.37)	1.425 (0.15)	1.426 (0.14)	1.389 (0.12)	1.335 (0.13)	1.352 (0.15)
Average Quarters Employed 3 Yrs before Entry	1.355 (0.21)	0.989 (0.34)	1.280 (0.14)	1.229 (0.12)	1.163 (0.12)	1.160 (0.14)
Proportion Never Worked in Yrs 1 to 3 before Entry	0.343 (0.06)	0.339 (0.05)	0.331 (0.05)	0.335 (0.04)	0.332 (0.05)	0.316 (0.04)
Average Real Earnings 1 Yr before Entry/\$1000	3.843 (0.54)	3.885 (0.46)	3.819 (0.51)	3.568 (0.44)	3.431 (0.48)	3.635 (0.52)
Average Real Earnings 2 Yrs before Entry/\$1000	3.692 (1.28)	4.819 (0.55)	4.563 (0.54)	4.135 (0.47)	3.743 (0.54)	3.747 (0.62)
Average Real Earnings 3 Yrs before Entry/\$1000	5.107 (1.00)	3.620 (1.35)	4.511 (0.60)	4.019 (0.51)	3.490 (0.54)	3.950 (2.65)
Number of Entrants	122,288	100,412	84,188	63,492	54,434	24,822

Notes:

Standard deviations between parentheses.

Averages and standard deviations are weighted by average county welfare caseload in the period.

/a First two quarters only

/b Infant=less than 1 year old

/c Toddler=1-5 years old

Table 3. Marginal Effects for proportions training types as functions of individual variables, local economic conditions and expenditures

Variable	Model (1)			Model (2)			Model (3)		
	NT	LFA	HCD	NT	LFA	HCD	NT	LFA	HCD
Log(Per Capita Expenditures)	-0.349*** (0.094)	0.278*** (0.070)	0.071** (0.032)	-0.288*** (0.102)	0.229*** (0.075)	0.058* (0.035)	-0.280*** (0.099)	0.225*** (0.073)	0.055 (0.034)
Log(Unemployment Rate)	0.106** (0.049)	-0.055 (0.037)	-0.051** (0.024)	0.108** (0.048)	-0.058 (0.036)	-0.051** (0.024)	0.100** (0.047)	-0.052 (0.036)	-0.048** (0.023)
Log(Employment Service Sector/Population)	-0.125* (0.071)	0.113** (0.050)	0.013 (0.032)	-0.133* (0.070)	0.121** (0.049)	0.013 (0.032)	-0.140** (0.069)	0.125*** (0.048)	0.015 (0.032)
Log(Average Earnings Service Sector)	-0.078 (0.059)	0.056 (0.044)	0.022 (0.034)	-0.089 (0.057)	0.065 (0.040)	0.024 (0.035)	-0.090 (0.058)	0.066 (0.041)	0.024 (0.034)
Proportion in 10th Decile Unobserv. Distrib. - 0.10	0.555* (0.288)	-0.694*** (0.219)	0.069 (0.131)	0.555** (0.266)	-1.110*** (0.211)	0.069 (0.126)	0.555 (0.716)	-0.555 (0.648)	0.069 (0.343)
Proportion in 90th Decile Unobserv. Distrib. - 0.10	1.110*** (0.345)	-1.804*** (0.264)	-0.278* (0.163)	1.110*** (0.346)	-0.833*** (0.263)	-0.555*** (0.159)	4.441*** (1.217)	-2.498*** (0.911)	-1.388*** (0.510)
Dummy CA TANF (CalWORKs) Period				-0.092*** (0.033)	0.074** (0.029)	0.019* (0.011)	-0.085*** (0.033)	0.070** (0.029)	0.015 (0.010)
Log(PC Budget) * Prop. 10th Decile Unobs. Dist.							0.000 (0.623)	0.000 (0.566)	0.000 (0.298)
Log(PC Budget) * Prop. 90th Decile Unobs. Dist.							2.220** (1.112)	-1.388 (0.855)	-1.318*** (0.445)
Observations	548	548	548	548	548	548	548	548	548
Controls for Demographic Characteristics	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
County Dummies	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Fiscal Year Dummies	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
F test Expenditures related terms = 0 (p-value)							0.011	0.012	0.005
F test Unobservables related terms = 0 (p-value)							0.000	0.000	0.060

Robust standard errors between parentheses (adjusted by county/fiscal year clusters)

* significant at 10%; ** significant at 5%; *** significant at 1%

Note:

NT = No Training (average = 0.66)

LFA = Labor Force Attachment Training (average = 0.25)

HCD = Human Capital Development Training (average = 0.09)

Results show marginal effects evaluated at mean values (except for dummy variables, where effect is change from 0 to 1), based on the results from running a multinomial logit model on the proportions of each type of training, see model results in Appendix Table A2

All models include average demographic characteristics of the cohorts of welfare entrants (gender, race, age, number and composition of children) as well as average labor market information before entry (employment and earnings each of the three years prior to welfare entry).

Table 4. Marginal Effects for proportions training types including political factors

Variable	Model (1)			Model (2)			Model (3)			Model (4)		
	NT	LFA	HCD	NT	LFA	HCD	NT	LFA	HCD	NT	LFA	HCD
Board of Supervisors Democratic (>50%)	-0.013 (0.029)	-0.016 (0.024)	0.030*** (0.011)	-0.009 (0.026)	-0.020 (0.023)	0.029*** (0.011)	-0.005 (0.026)	-0.024 (0.023)	0.028** (0.012)			
Board of Supervisors 'Strong' Democratic (≥80%)										-0.045 (0.055)	-0.016 (0.042)	0.062* (0.032)
Board of Supervisors 'Weak' Democratic (60%)										-0.008 (0.026)	-0.020 (0.023)	0.028** (0.011)
Log(Per Capita Expenditures)	-0.347*** (0.095)	0.276*** (0.070)	0.070** (0.032)	-0.286*** (0.102)	0.227*** (0.076)	0.059* (0.035)	-0.289*** (0.102)	0.229*** (0.076)	0.059* (0.035)	-0.289*** (0.103)	0.229*** (0.077)	0.060* (0.035)
Log(Unemployment Rate)	0.107** (0.049)	-0.056 (0.038)	-0.050** (0.024)	0.109** (0.048)	-0.059 (0.037)	-0.050** (0.024)	0.107** (0.050)	-0.058 (0.037)	-0.049** (0.025)	0.114** (0.049)	-0.060* (0.036)	-0.054** (0.025)
Log(Employment Service Sector/Population)	-0.128* (0.073)	0.110** (0.051)	0.018 (0.032)	-0.135* (0.071)	0.118** (0.050)	0.018 (0.032)	-0.133* (0.070)	0.116** (0.049)	0.017 (0.031)	-0.135* (0.071)	0.117** (0.050)	0.018 (0.032)
Log(Average Earnings Service Sector)	-0.080 (0.060)	0.057 (0.044)	0.023 (0.034)	-0.091 (0.058)	0.066 (0.041)	0.025 (0.035)	-0.109* (0.058)	0.078** (0.039)	0.031 (0.035)	-0.091 (0.058)	0.065 (0.041)	0.026 (0.034)
Proportion in 10th Decile Unobserv. Distrib. - 0.10	0.000 (0.289)	0.000 (0.222)	0.000 (0.129)	0.000 (0.268)	0.000 (0.211)	0.000 (0.125)	0.000 (0.313)	0.000 (0.235)	0.000 (0.148)	0.555** (0.269)	-0.833*** (0.214)	0.069 (0.125)
Proportion in 90th Decile Unobserv. Distrib. - 0.10	1.665*** (0.343)	-1.527*** (0.263)	-0.486*** (0.160)	1.665*** (0.347)	-1.527*** (0.264)	-0.139 (0.158)	1.665*** (0.404)	-0.694** (0.292)	-0.208 (0.181)	1.110*** (0.338)	-0.694*** (0.260)	-0.555*** (0.153)
Dummy CA TANF (CalWORKs) Period				-0.092*** (0.033)	0.075** (0.029)	0.017 (0.011)	-0.094*** (0.033)	0.076*** (0.029)	0.018 (0.011)	-0.092*** (0.033)	0.074** (0.029)	0.018 (0.011)
Democratic Board * Prop. 10th Decile Unobs. Dist.							-1.665*** (0.432)	0.694** (0.347)	0.208 (0.243)			
Democratic Board * Prop. 90th Decile Unobs. Dist.							1.340** (0.641)	-0.909* (0.526)	-0.432 (0.311)			
Observations	548	548	548	548	548	548	548	548	548	548	548	548
Controls for Demographic Characteristics	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
County Dummies	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Fiscal Year Dummies	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
F test Democratic Board related terms = 0 (p-value)							0.000	0.046	0.032	0.704	0.669	0.038
F test Unobservables related terms = 0 (p-value)							0.000	0.000	0.097			

Robust standard errors between parentheses (adjusted by county/fiscal year clusters)

* significant at 10%; ** significant at 5%; *** significant at 1%

Note:

NT = No Training (average = 0.66)

LFA = Labor Force Attachment Training (average = 0.25)

HCD = Human Capital Development Training (average = 0.09)

Results show marginal effects evaluated at mean values (except for dummy variables, where effect is change from 0 to 1), based on the results from running a multinomial logit model on the proportions of each type of training, see model results in Appendix Table A3

All models include average demographic characteristics of the cohorts of welfare entrants (gender, race, age, number and composition of children) as well as average labor market information before entry (employment and earnings each of the three years prior to welfare entry).

Table 5. Marginal Effects for proportions training types, HCD training components

Variable	Model (1)				Model (2)				Model (3)			
	NT	LFA	ABE	VOC	NT	LFA	ABE	VOC	NT	LFA	ABE	VOC
Board of Supervisors Democratic (>50%)	-0.015 (0.029)	-0.016 (0.025)	0.015*** (0.006)	0.015** (0.006)	-0.010 (0.027)	-0.020 (0.023)	0.016*** (0.006)	0.014** (0.006)				
Board of Supervisors 'Strong' Democratic (≥80%)									-0.046 (0.054)	-0.013 (0.043)	0.017 (0.013)	0.043** (0.020)
Board of Supervisors 'Weak' Democratic (60%)									-0.009 (0.026)	-0.020 (0.023)	0.015*** (0.006)	0.014** (0.006)
Log(Per Capita Expenditures)	-0.340*** (0.092)	0.280*** (0.071)	0.009 (0.020)	0.051*** (0.013)	-0.283*** (0.100)	0.230*** (0.077)	0.009 (0.022)	0.044*** (0.014)	-0.285*** (0.101)	0.232*** (0.077)	0.009 (0.021)	0.045*** (0.014)
Log(Unemployment Rate)	0.098** (0.048)	-0.062* (0.038)	-0.004 (0.015)	-0.031*** (0.010)	0.100** (0.047)	-0.065* (0.037)	-0.003 (0.015)	-0.031*** (0.010)	0.105** (0.047)	-0.067* (0.036)	-0.004 (0.015)	-0.035*** (0.010)
Log(Employment Service Sector/Population)	-0.131* (0.071)	0.109** (0.052)	0.010 (0.018)	0.012 (0.012)	-0.139** (0.069)	0.117** (0.051)	0.011 (0.018)	0.012 (0.012)	-0.140** (0.069)	0.116** (0.051)	0.011 (0.018)	0.013 (0.012)
Log(Average Earnings Service Sector)	-0.077 (0.056)	0.056 (0.045)	0.038* (0.022)	-0.017 (0.014)	-0.087 (0.054)	0.064 (0.041)	0.039* (0.022)	-0.015 (0.013)	-0.087 (0.054)	0.063 (0.041)	0.039* (0.022)	-0.015 (0.013)
Proportion in 10th Decile Unobserv. Distrib. - 0.10	0.555** (0.280)	-0.555** (0.223)	0.000 (0.071)	0.035 (0.069)	0.555** (0.259)	-0.833*** (0.212)	0.243*** (0.070)	0.035 (0.069)	0.000 (0.263)	0.000 (0.217)	0.173** (0.069)	0.000 (0.067)
Proportion in 90th Decile Unobserv. Distrib. - 0.10	0.555* (0.331)	-0.555** (0.268)	-0.173** (0.083)	-0.451*** (0.087)	0.555* (0.334)	-0.833*** (0.267)	-0.173** (0.080)	-0.312*** (0.083)	1.665*** (0.329)	-1.527*** (0.266)	-0.069 (0.082)	-0.208*** (0.079)
Dummy CA TANF (CalWORKs) Period					-0.086*** (0.033)	0.076** (0.029)	0.000 (0.006)	0.010** (0.005)	-0.086*** (0.033)	0.075** (0.029)	0.000 (0.006)	0.011** (0.005)
Observations	548	548	548	548	548	548	548	548	548	548	548	548
Controls for Demographic Characteristics	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
County Dummies	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Fiscal Year Dummies	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
F test Democratic Board related terms = 0 (p-value)									0.690	0.685	0.033	0.036

Robust standard errors between parentheses (adjusted by county/fiscal year clusters)

* significant at 10%; ** significant at 5%; *** significant at 1%

Note:

NT = No Training (average = 0.66)

LFA = Labor Force Attachment Training (average = 0.25)

HCD = Human Capital Development Training (average = 0.09)

ABE = Adult Basic Education Training (average = 0.05)

VOC = Vocational Training (average = 0.04)

Results show marginal effects evaluated at mean values (except for dummy variables, where effect is change from 0 to 1), based on the results from running a multinomial logit model on the proportions of each type of training, see model results in Appendix Table A4

All models include average demographic characteristics of the cohorts of welfare entrants (gender, race, age, number and composition of children) as well as average labor market information before entry (employment and earnings each of the three years prior to welfare entry).

Figure 1
Proportion Training Types by Year

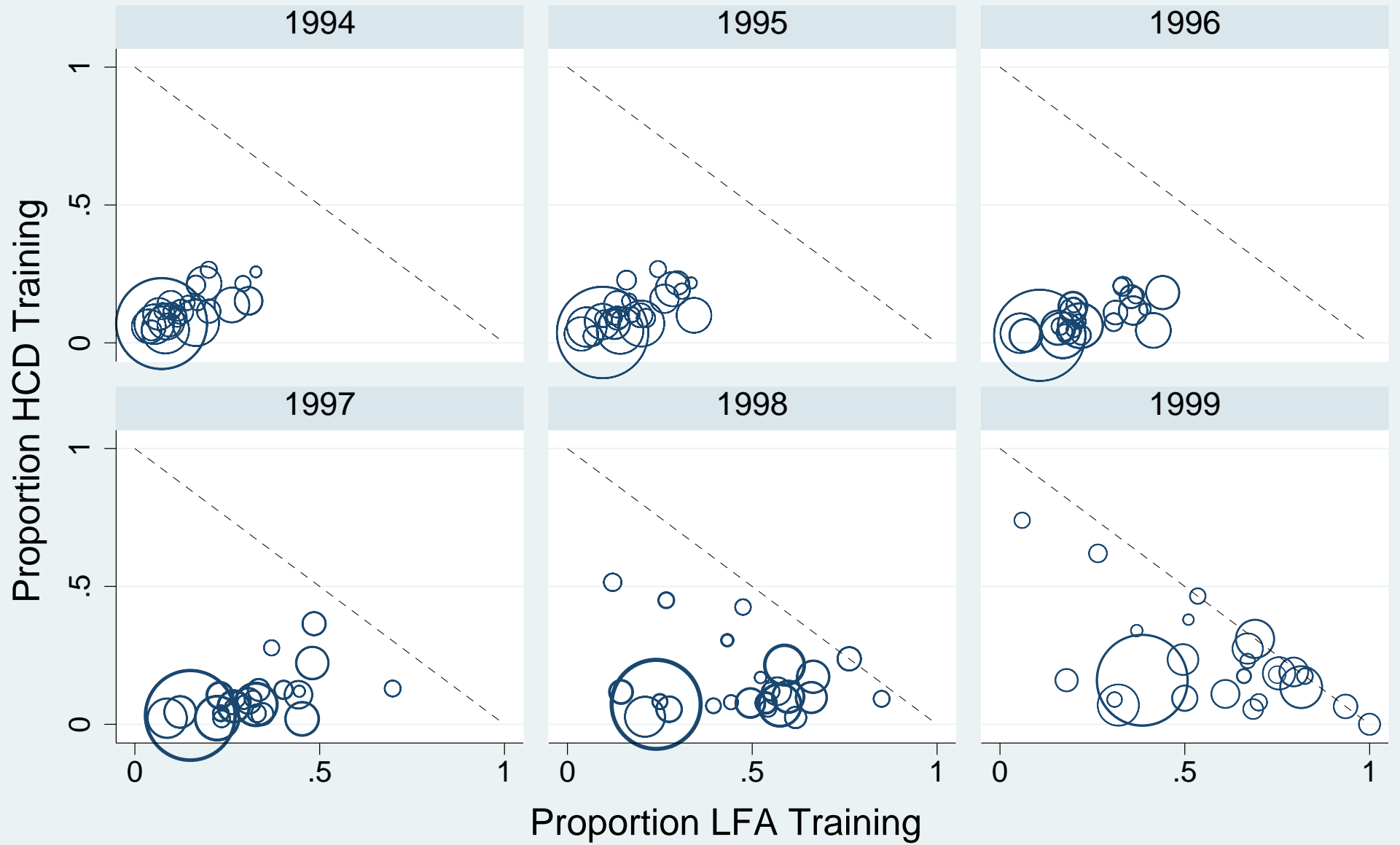


Figure 2. Interior Solution Cases for (P1)

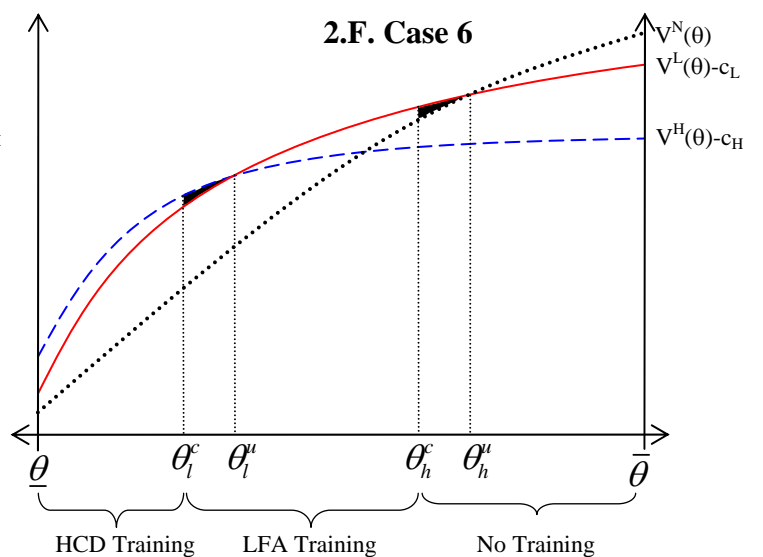
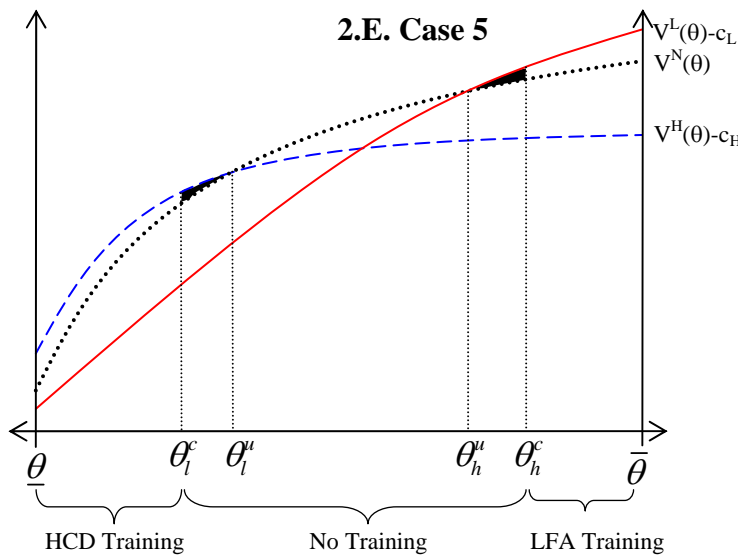
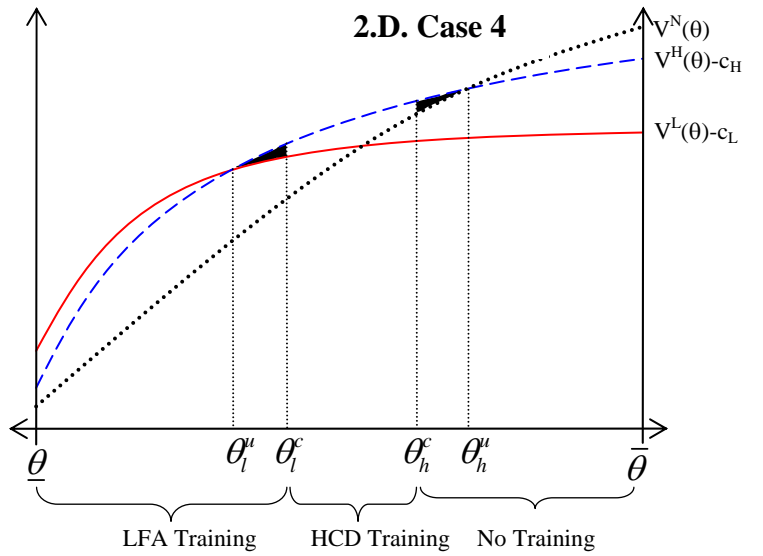
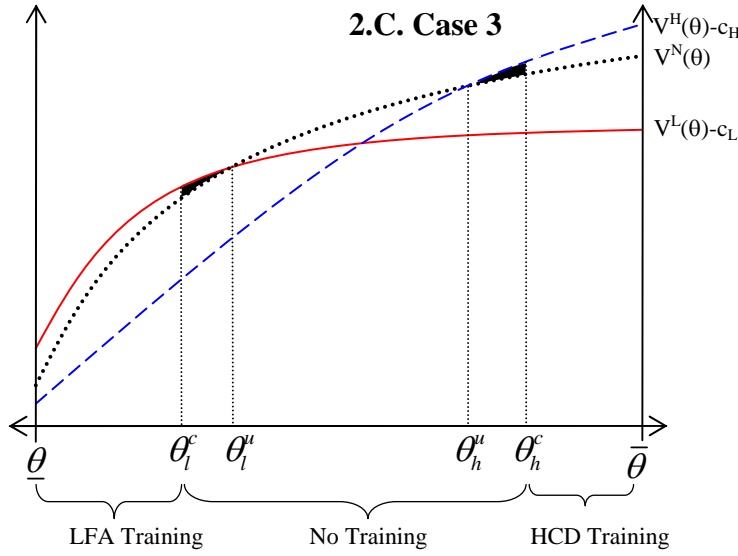
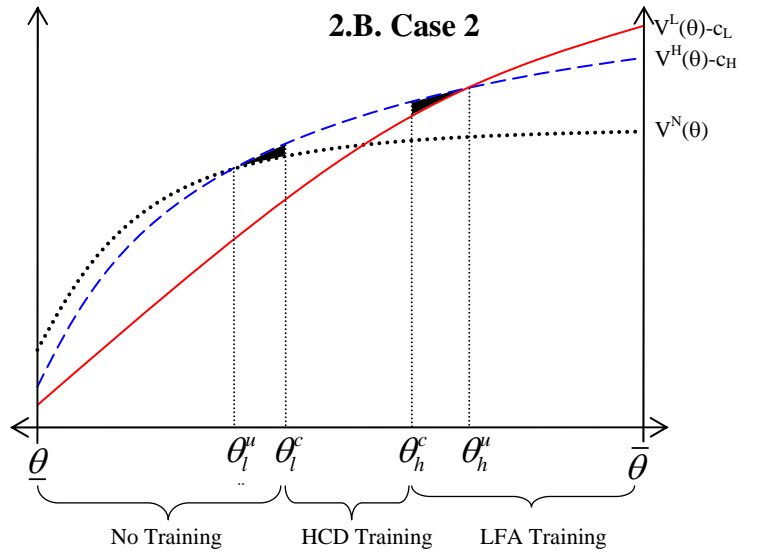
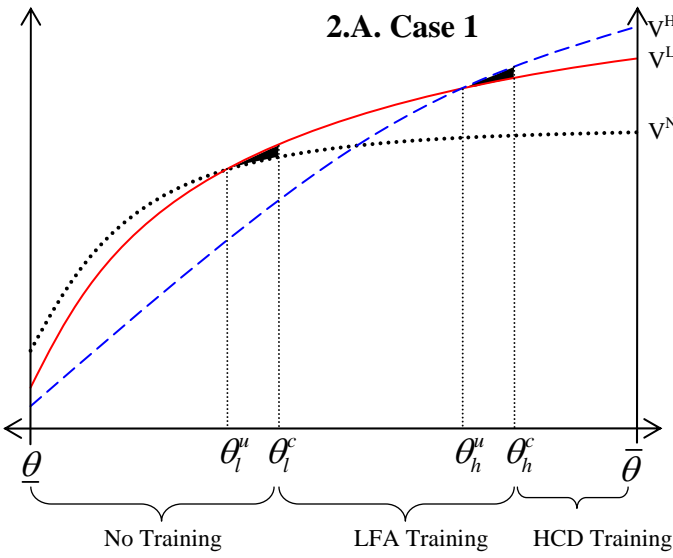


Figure 3

Deciles Comparison of Estimated Unobservables

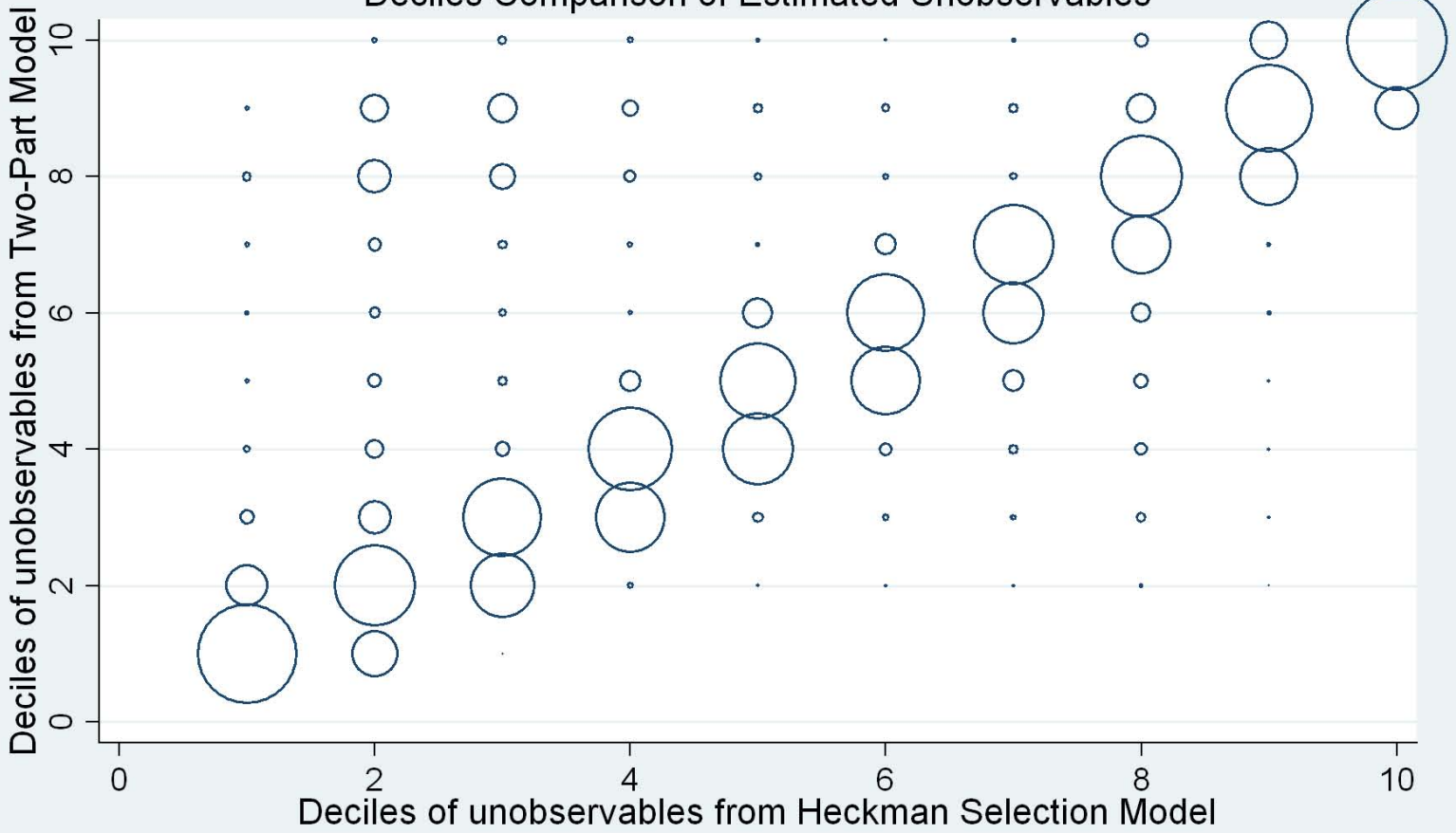
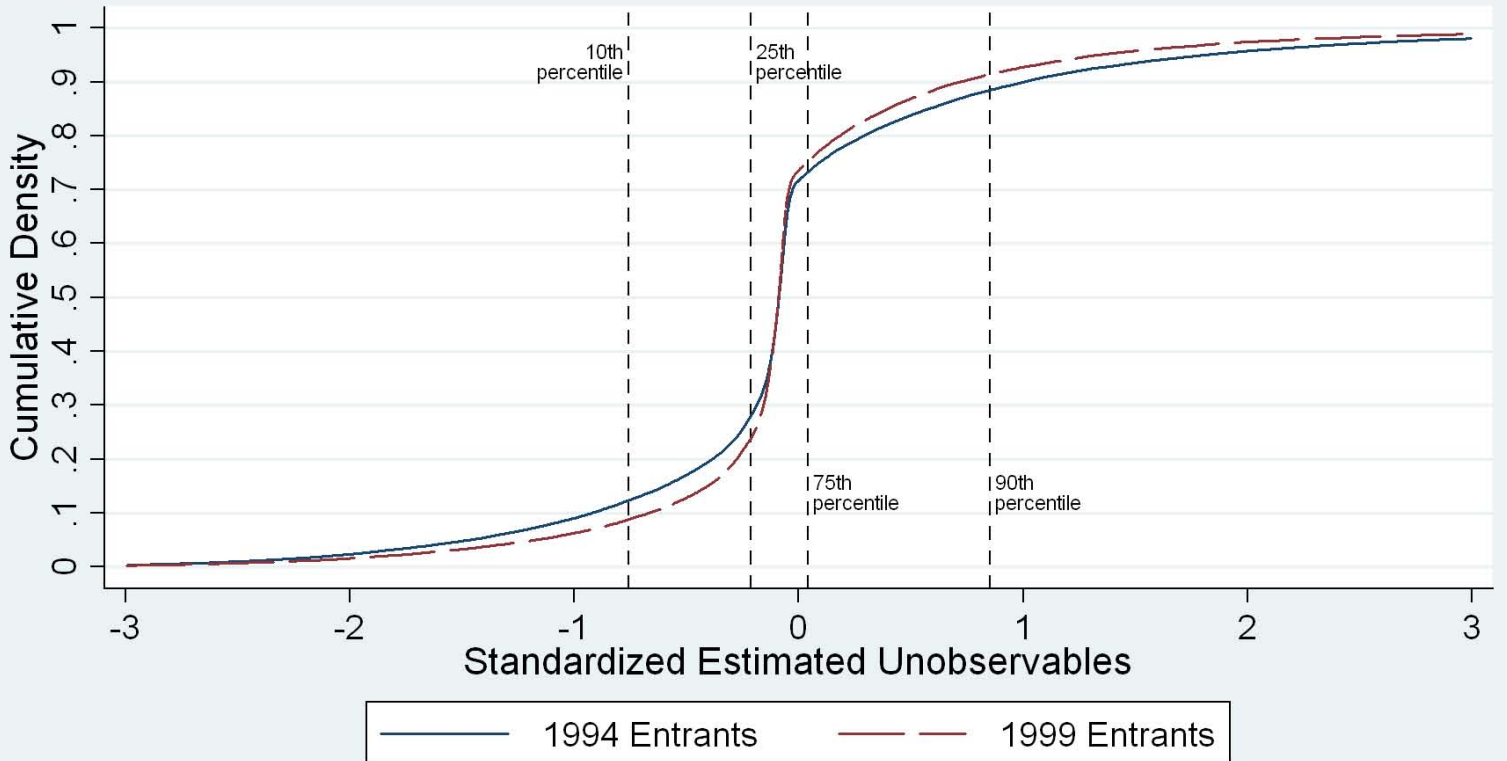


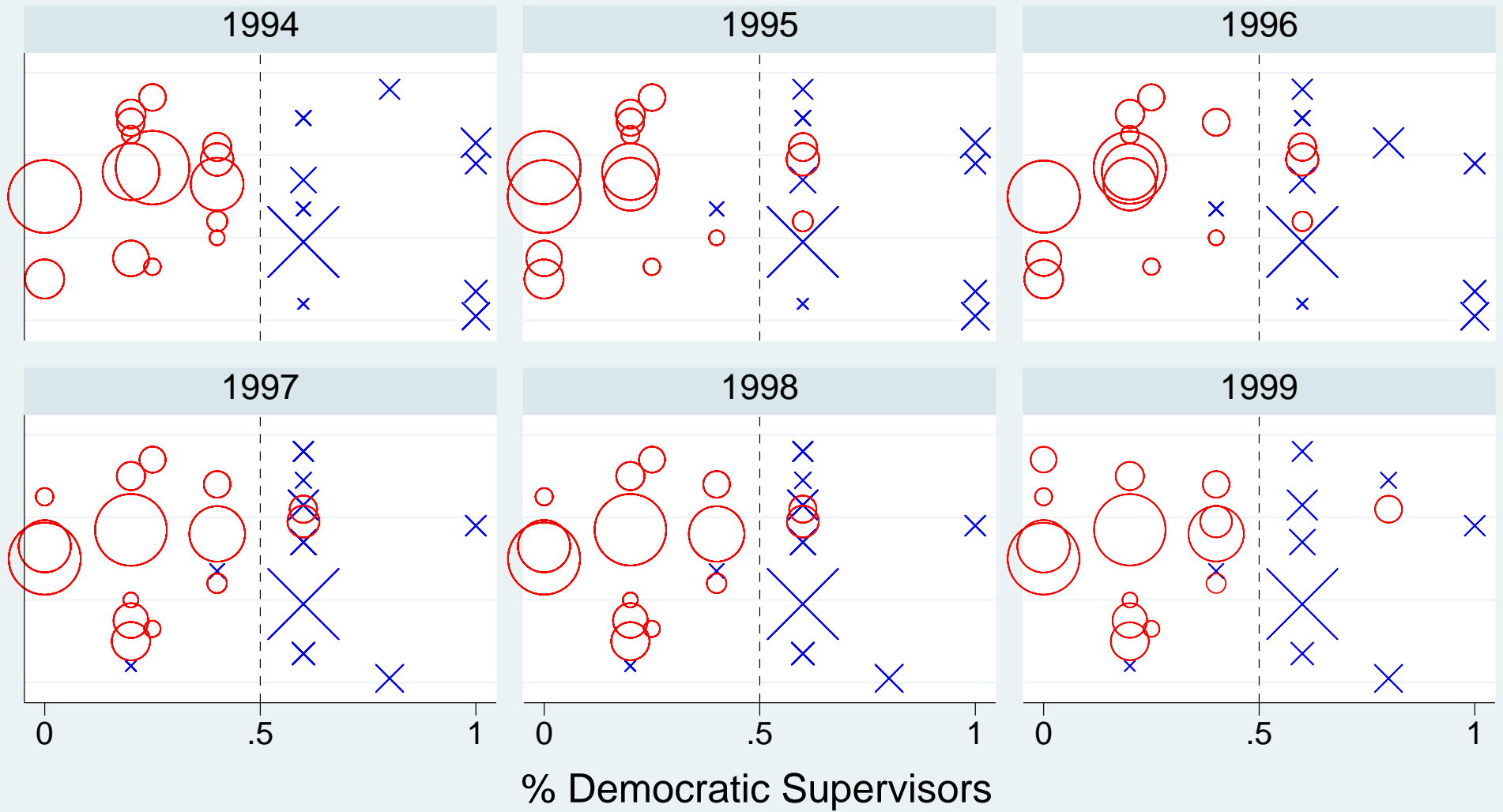
Figure 4

CDFs of Estimated Unobservables



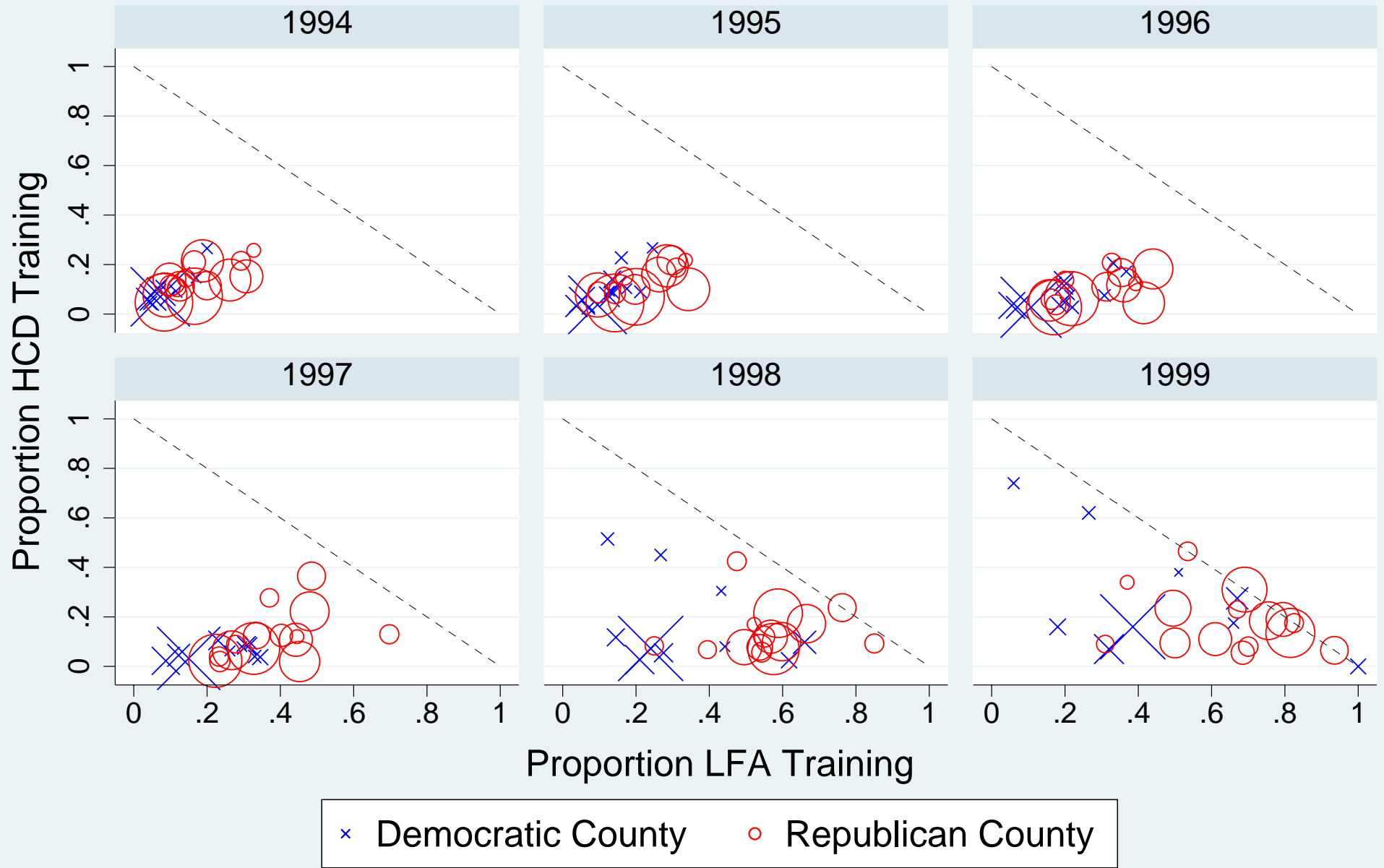
Note: Vertical lines mark percentiles from the pooled distribution, not from particular years

Figure 5
County Board of Supervisors Composition



× Democratic Majority in 1994 ○ Republican Majority in 1994

Figure 6
Proportion Training Types by Year and Party



Appendix Table A1. Estimation of unobservables by Heckman Selection Model and by Two-Part Model

Variable	Heckman Selection Model		Two-Part Model	
	Employment	log(Earnings)	Employment	Earnings
Age	-0.024** (0.004)	0.240** (0.004)	-0.038** (0.010)	0.215** (0.005)
Age squared	0.000 (0.000)	-0.003** (0.000)	0.000 (0.000)	-0.003** (0.000)
Female	-0.342** (0.008)	-0.271** (0.008)	-0.626** (0.021)	-0.247** (0.009)
White	-0.090** (0.011)	-0.050** (0.014)	-0.095** (0.029)	-0.004 (0.018)
Hispanic	0.059** (0.012)	0.038** (0.015)	0.131** (0.031)	0.021 (0.019)
Black	-0.079** (0.013)	-0.133** (0.016)	-0.096** (0.033)	-0.031 (0.021)
Language: English	0.615** (0.015)	0.182** (0.024)	1.040** (0.035)	0.139** (0.025)
Language: Spanish	0.654** (0.018)	0.007 (0.026)	1.079** (0.045)	-0.161** (0.028)
Single Parent Family	0.014 (0.008)		0.020 (0.020)	0.006 (0.009)
Received AFDC as a child	-0.076** (0.007)		-0.164** (0.020)	-0.129** (0.012)
Number of kids in household	-0.060** (0.005)		-0.106** (0.014)	0.006 (0.006)
Dummy infant in household	-0.040** (0.008)		-0.072** (0.021)	-0.024* (0.010)
Dummy kids 1-5 yrs old in household	-0.165** (0.008)		-0.314** (0.022)	0.077** (0.011)
Dummy kids 6-12 yrs old in household	-0.028* (0.011)		-0.092** (0.031)	0.038** (0.014)
Dummy employed one quarter before this period	0.384** (0.009)	1.011** (0.021)	0.636** (0.026)	0.623** (0.010)
Dummy employed two quarters before this period	0.345** (0.010)	0.335** (0.011)	2.472** (0.025)	0.250** (0.009)
Dummy employed three quarters before this period	1.397** (0.009)	0.282** (0.011)	0.607** (0.028)	0.194** (0.010)
Dummy employed four quarters before this period	0.431** (0.010)	0.370** (0.010)	0.755** (0.029)	0.150** (0.010)
Growth Rate Total Employment/Population	0.083 (0.190)	-0.364 (0.304)	0.868 (0.263)	-0.353* (0.252)
Growth Rate Empl. Retail Sector/Population	0.227 (0.276)	0.169 (0.230)	-0.201 (0.353)	0.159 (0.053)
Growth Rate Empl. Service Sector/Population	-0.162 (0.343)	0.032 (0.058)	0.865 (0.510)	-0.620* (0.227)
Growth Rate Total Average Real Earnings	-0.056 (0.131)	-0.560 (0.384)	-0.221* (0.446)	0.045 (0.165)
Growth Rate Retail Sector Average Real Earnings	-0.117 (0.052)	-0.516* (0.204)	-0.128 (0.098)	0.040 (0.345)
Growth Rate Service Sector Average Real Earnings	0.128 (0.206)	0.004 (0.147)	1.041 (0.679)	0.072 (0.135)
Constant	-0.361** (0.058)	-3.834** (0.068)	-0.638** (0.153)	-2.298** (0.086)
Inverse Mills ratio (λ)	0.310 (0.029)			
ρ	0.248			
σ	1.249			
Observations	305,120	156,271	305,120	156,271

* significant at 5%; ** significant at 1%
Robust standard errors in parentheses

Notes:

All regressions include also county dummies and year/quarter dummies.

The employment and earnings dependent variables refer to two years (quarters 5-8) before first welfare entry.

The first two columns show the results from estimating the two-step Heckman selection model, using log(earnings).

The second two columns show the results from estimating a two-part model with a logit regression for the employment equation, and a GLM model with a Gamma function and log link, for the earnings equation (in levels)

Appendix Table A2. Multinomial Logit Model for Training Proportion (basis for Table 3)

Variable	Model (1)		Model (2)		Model (3)	
	LFA	HCD	LFA	HCD	LFA	HCD
Log(Per Capita Expenditures)	1.886*** (0.468)	1.382*** (0.516)	1.559*** (0.511)	1.137** (0.564)	1.524*** (0.494)	1.077** (0.542)
Log(Unemployment Rate)	-0.426* (0.251)	-0.792** (0.359)	-0.441* (0.245)	-0.791** (0.362)	-0.403* (0.241)	-0.739** (0.346)
Log(Employment Service Sector/Population)	0.742** (0.344)	0.333 (0.489)	0.795** (0.337)	0.343 (0.485)	0.824** (0.333)	0.385 (0.483)
Log(Average Earnings Service Sector)	0.390 (0.294)	0.380 (0.491)	0.451* (0.273)	0.425 (0.498)	0.459* (0.279)	0.431 (0.495)
Proportion in 10th Decile Unobserv. Distrib. - 0.10	-2.690* (1.488)	1.072 (1.950)	-1.159 (1.411)	2.006 (1.848)	-4.372 (4.168)	-1.029 (4.836)
Proportion in 90th Decile Unobserv. Distrib. - 0.10	-8.849*** (1.767)	-7.925*** (2.436)	-7.247*** (1.769)	-6.895*** (2.390)	-18.014*** (6.165)	-24.577*** (7.711)
Proportion Female	-0.435 (0.987)	0.263 (1.343)	1.461 (1.183)	1.565 (1.488)	1.628 (1.185)	1.859 (1.449)
Proportion White	5.357** (2.357)	2.063 (3.153)	6.643*** (2.288)	2.950 (3.110)	5.512*** (1.918)	0.641 (2.871)
Proportion Hispanic	3.771* (2.115)	2.579 (2.839)	5.137** (2.119)	3.561 (2.768)	4.215** (1.923)	1.607 (2.600)
Proportion Black	7.634** (3.020)	-0.101 (4.652)	8.422*** (2.938)	0.410 (4.562)	7.018*** (2.633)	-2.191 (4.347)
Average Age	0.173** (0.083)	0.015 (0.101)	0.214*** (0.082)	0.043 (0.101)	0.204** (0.081)	0.030 (0.104)
Number of Children	-1.406** (0.572)	-0.775 (0.633)	-1.391** (0.550)	-0.773 (0.627)	-1.414*** (0.540)	-0.765 (0.614)
Proportion with Children < 1 year old	3.092** (1.339)	1.180 (1.625)	2.738** (1.362)	0.860 (1.678)	2.610** (1.273)	0.684 (1.644)
Proportion with Children 1-5 years old	-0.787 (1.229)	0.384 (1.499)	-0.611 (1.214)	0.474 (1.500)	-0.607 (1.184)	0.467 (1.514)
Avg. Employment Year 1 before Welfare Entry	0.362 (0.586)	0.204 (0.822)	0.503 (0.585)	0.265 (0.813)	0.587 (0.598)	0.366 (0.822)
Avg. Employment Year 2 before Welfare Entry	-0.040 (0.398)	-0.603 (0.576)	-0.052 (0.383)	-0.632 (0.570)	-0.157 (0.392)	-0.756 (0.584)
Avg. Employment Year 3 before Welfare Entry	-0.385 (0.257)	-0.597** (0.241)	-0.422 (0.258)	-0.638*** (0.240)	-0.417 (0.255)	-0.650*** (0.240)
Prop. Never Employed Yrs 1-3 before Welfare Entry	-1.275 (0.959)	-4.513*** (1.553)	-0.600 (0.972)	-4.226*** (1.522)	-0.598 (0.975)	-4.237*** (1.512)
Avg. Earnings Year 1 before Welfare Entry/\$1,000	0.072 (0.138)	0.139 (0.188)	0.060 (0.138)	0.129 (0.185)	0.069 (0.137)	0.139 (0.178)
Avg. Earnings Year 2 before Welfare Entry/\$1,000	-0.003 (0.106)	0.069 (0.150)	0.010 (0.103)	0.082 (0.149)	0.047 (0.107)	0.131 (0.153)
Avg. Earnings Year 3 before Welfare Entry/\$1,000	0.093 (0.061)	0.130*** (0.047)	0.099* (0.060)	0.137*** (0.047)	0.090 (0.059)	0.129*** (0.047)
Dummy CA TANF (CalWORKs) Period			0.470*** (0.170)	0.356** (0.160)	0.444*** (0.169)	0.301** (0.148)
Log(PC Budget) * Prop. 10th Decile Unobs. Dist.					-3.475 (3.639)	-3.280 (4.206)
Log(PC Budget) * Prop. 90th Decile Unobs. Dist.					-10.643* (5.765)	-17.563*** (6.704)
Constant	-17.187*** (4.765)	-4.081 (7.439)	-22.403*** (5.347)	-7.202 (7.742)	-21.791*** (5.217)	-5.887 (7.648)
Observations	548	548	548	548	548	548
Pseudo R-squared	0.136		0.137		0.137	
County Dummies	Yes	Yes	Yes	Yes	Yes	Yes
Fiscal Year Dummies	Yes	Yes	Yes	Yes	Yes	Yes

Robust standard errors between parentheses (adjusted by county/fiscal year clusters)

* significant at 10%; ** significant at 5%; *** significant at 1%

Appendix Table A3. Multinomial Model Estimation (basis for Table 4)

Variable	Model (1)		Model (2)		Model (3)		Model (4)	
	LFA	HCD	LFA	HCD	LFA	HCD	LFA	HCD
Board of Supervisors Democratic (>50%)	-0.064 (0.158)	0.404** (0.176)	-0.090 (0.149)	0.393** (0.169)	-0.111 (0.148)	0.378** (0.174)		
Board of Supervisors 'Strong' Democratic (≥80%)							-0.017 (0.285)	0.765** (0.330)
Board of Supervisors 'Weak' Democratic (60%)							-0.092 (0.149)	0.393** (0.172)
Log(Per Capita Expenditures)	1.874*** (0.471)	1.371*** (0.512)	1.545*** (0.515)	1.139** (0.557)	1.560*** (0.513)	1.150** (0.558)	1.558*** (0.518)	1.165** (0.560)
Log(Unemployment Rate)	-0.430* (0.252)	-0.787** (0.357)	-0.447* (0.247)	-0.784** (0.360)	-0.439* (0.251)	-0.771** (0.370)	-0.460* (0.245)	-0.841** (0.365)
Log(Employment Service Sector/Population)	0.732** (0.353)	0.400 (0.491)	0.782** (0.345)	0.410 (0.487)	0.769** (0.338)	0.404 (0.477)	0.781** (0.345)	0.416 (0.485)
Log(Average Earnings Service Sector)	0.398 (0.297)	0.403 (0.494)	0.458* (0.276)	0.447 (0.501)	0.545** (0.265)	0.540 (0.511)	0.452 (0.276)	0.455 (0.496)
Proportion in 10th Decile Unobserv. Distrib. - 0.10	-2.676* (1.489)	1.233 (1.935)	-1.135 (1.409)	2.126 (1.844)	-3.237** (1.588)	-0.109 (2.197)	-1.201 (1.432)	2.024 (1.842)
Proportion in 90th Decile Unobserv. Distrib. - 0.10	-8.886*** (1.765)	-7.740*** (2.409)	-7.280*** (1.769)	-6.748*** (2.364)	-4.697** (1.993)	-3.639 (2.747)	-7.231*** (1.748)	-6.676*** (2.312)
Proportion Female	-0.451 (0.989)	0.362 (1.343)	1.455 (1.184)	1.616 (1.480)	1.503 (1.179)	1.683 (1.476)	1.462 (1.180)	1.624 (1.471)
Proportion White	5.421** (2.350)	1.897 (3.182)	6.722*** (2.278)	2.749 (3.135)	6.546*** (2.189)	2.518 (3.059)	6.708*** (2.210)	2.451 (2.927)
Proportion Hispanic	3.810* (2.107)	2.528 (2.849)	5.185** (2.112)	3.463 (2.776)	5.141** (2.042)	3.374 (2.725)	5.181** (2.068)	3.227 (2.620)
Proportion Black	7.675** (3.014)	-0.220 (4.680)	8.475*** (2.927)	0.271 (4.591)	8.466*** (2.881)	0.245 (4.517)	8.379*** (2.829)	-0.314 (4.351)
Average Age	0.166** (0.083)	0.023 (0.103)	0.207** (0.082)	0.049 (0.103)	0.215*** (0.080)	0.061 (0.102)	0.209** (0.083)	0.061 (0.103)
Number of Children	-1.353** (0.568)	-0.819 (0.648)	-1.338** (0.544)	-0.816 (0.643)	-1.285** (0.536)	-0.774 (0.636)	-1.332** (0.540)	-0.825 (0.641)
Proportion with Children < 1 year old	3.049** (1.341)	1.269 (1.624)	2.693** (1.365)	0.952 (1.675)	2.488* (1.387)	0.764 (1.693)	2.675** (1.364)	0.932 (1.675)
Proportion with Children 1-5 years old	-0.882 (1.228)	0.610 (1.512)	-0.712 (1.211)	0.692 (1.513)	-0.971 (1.204)	0.450 (1.538)	-0.750 (1.211)	0.627 (1.523)
Avg. Employment Year 1 before Welfare Entry	0.373 (0.586)	0.209 (0.815)	0.514 (0.585)	0.270 (0.807)	0.315 (0.585)	0.049 (0.788)	0.492 (0.584)	0.191 (0.798)
Avg. Employment Year 2 before Welfare Entry	-0.020 (0.399)	-0.665 (0.564)	-0.028 (0.383)	-0.693 (0.558)	0.104 (0.386)	-0.525 (0.579)	-0.013 (0.382)	-0.630 (0.559)
Avg. Employment Year 3 before Welfare Entry	-0.382 (0.257)	-0.602** (0.239)	-0.419 (0.258)	-0.643*** (0.238)	-0.381 (0.259)	-0.602** (0.240)	-0.405 (0.256)	-0.623*** (0.236)
Prop. Never Employed Yrs 1-3 before Welfare Entry	-1.210 (0.943)	-4.641*** (1.544)	-0.524 (0.952)	-4.363*** (1.516)	-0.685 (0.930)	-4.526*** (1.513)	-0.541 (0.953)	-4.417*** (1.530)
Avg. Earnings Year 1 before Welfare Entry/\$1,000	0.071 (0.138)	0.126 (0.188)	0.059 (0.137)	0.116 (0.185)	0.083 (0.137)	0.143 (0.179)	0.063 (0.136)	0.132 (0.181)
Avg. Earnings Year 2 before Welfare Entry/\$1,000	-0.007 (0.106)	0.082 (0.148)	0.006 (0.103)	0.095 (0.147)	-0.018 (0.104)	0.063 (0.151)	0.002 (0.103)	0.082 (0.148)
Avg. Earnings Year 3 before Welfare Entry/\$1,000	0.094 (0.061)	0.129*** (0.047)	0.100* (0.060)	0.136*** (0.047)	0.098 (0.061)	0.134*** (0.047)	0.096 (0.060)	0.131*** (0.046)
Dummy CA TANF (CalWORKs) Period			0.474*** (0.170)	0.340** (0.156)	0.484*** (0.167)	0.347** (0.158)	0.473*** (0.170)	0.344** (0.156)
Democratic Board * Prop. 10th Decile Unobs. Dist.					4.692** (2.264)	4.771 (3.462)		
Democratic Board * Prop. 90th Decile Unobs. Dist.					-6.452* (3.479)	-7.333 (4.550)		
Constant	-16.926*** (4.953)	-5.499 (7.559)	-22.102*** (5.502)	-8.484 (7.831)	-22.035*** (5.501)	-8.581 (7.769)	-22.101*** (5.562)	-8.682 (7.859)
Observations	548	548	548	548	548	548	548	548
Pseudo R-squared	0.136		0.137		0.137		0.137	
County Dummies	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Fiscal Year Dummies	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Robust standard errors between parentheses (adjusted by county/fiscal year clusters)

* significant at 10%; ** significant at 5%; *** significant at 1%

Appendix Table A4. Multinomial Model Estimation (basis for Table 5)

Variable	Model (1) HCD			Model (2) HCD			Model (3) HCD		
	LFA	ABE	VOC	LFA	ABE	VOC	LFA	ABE	VOC
Board of Supervisors Democratic (>50%)	-0.058 (0.161)	0.404** (0.163)	0.494** (0.215)	-0.086 (0.150)	0.401** (0.160)	0.469** (0.205)			
Board of Supervisors 'Strong' Democratic (≥80%)							-0.001 (0.286)	0.478 (0.309)	1.115*** (0.380)
Board of Supervisors 'Weak' Democratic (60%)							-0.087 (0.150)	0.400** (0.160)	0.482** (0.209)
Log(Per Capita Expenditures)	1.869*** (0.470)	0.692 (0.567)	2.051*** (0.490)	1.544*** (0.513)	0.602 (0.625)	1.762*** (0.534)	1.557*** (0.516)	0.612 (0.624)	1.781*** (0.538)
Log(Unemployment Rate)	-0.447* (0.251)	-0.233 (0.395)	-1.102*** (0.353)	-0.466* (0.245)	-0.218 (0.404)	-1.105*** (0.353)	-0.480** (0.242)	-0.231 (0.408)	-1.216*** (0.355)
Log(Employment Service Sector/Population)	0.727** (0.353)	0.422 (0.517)	0.536 (0.458)	0.777** (0.345)	0.445 (0.509)	0.551 (0.452)	0.778** (0.344)	0.447 (0.508)	0.596 (0.449)
Log(Average Earnings Service Sector)	0.384 (0.293)	1.024* (0.579)	-0.417 (0.447)	0.442 (0.272)	1.062* (0.583)	-0.357 (0.444)	0.436 (0.273)	1.058* (0.583)	-0.338 (0.432)
Proportion in 10th Decile Unobserv. Distrib. - 0.10	-2.658* (1.486)	1.872 (1.955)	0.339 (2.357)	-1.125 (1.405)	2.308 (1.889)	1.680 (2.304)	-1.198 (1.428)	2.263 (1.890)	1.585 (2.265)
Proportion in 90th Decile Unobserv. Distrib. - 0.10	-8.940*** (1.771)	-4.962** (2.336)	-11.155*** (2.882)	-7.343*** (1.773)	-4.516** (2.299)	-9.740*** (2.781)	-7.282*** (1.752)	-4.477* (2.286)	-9.779*** (2.688)
Proportion Female	-0.428 (0.995)	2.048 (1.321)	-1.353 (1.540)	1.458 (1.184)	2.652* (1.436)	0.292 (1.678)	1.468 (1.179)	2.652* (1.433)	0.300 (1.656)
Proportion White	5.334** (2.352)	3.958* (2.388)	0.029 (4.459)	6.624*** (2.276)	4.275* (2.379)	1.216 (4.337)	6.600*** (2.203)	4.258* (2.324)	0.536 (4.054)
Proportion Hispanic	3.777* (2.121)	5.194** (2.256)	0.400 (4.108)	5.137** (2.123)	5.511** (2.260)	1.703 (3.965)	5.120** (2.073)	5.500** (2.219)	1.225 (3.748)
Proportion Black	7.645** (3.026)	-0.317 (3.981)	-1.425 (6.574)	8.447*** (2.936)	-0.170 (3.969)	-0.734 (6.466)	8.329*** (2.838)	-0.263 (3.812)	-1.763 (6.221)
Average Age	0.169** (0.083)	-0.090 (0.106)	0.076 (0.110)	0.210*** (0.081)	-0.082 (0.105)	0.110 (0.108)	0.212** (0.083)	-0.080 (0.105)	0.138 (0.110)
Number of Children	-1.373** (0.571)	-0.559 (0.717)	-1.206* (0.703)	-1.358** (0.546)	-0.548 (0.715)	-1.197* (0.688)	-1.350** (0.541)	-0.545 (0.715)	-1.263* (0.684)
Proportion with Children < 1 year old	3.089** (1.350)	1.018 (1.702)	1.193 (2.034)	2.734** (1.376)	0.859 (1.737)	0.856 (2.110)	2.716** (1.375)	0.846 (1.736)	0.839 (2.107)
Proportion with Children 1-5 years old	-0.837 (1.225)	0.675 (1.547)	0.988 (1.770)	-0.664 (1.207)	0.691 (1.558)	1.084 (1.756)	-0.702 (1.207)	0.658 (1.564)	1.095 (1.762)
Avg. Employment Year 1 before Welfare Entry	0.386 (0.584)	-0.784 (0.883)	0.760 (0.814)	0.529 (0.582)	-0.774 (0.878)	0.854 (0.810)	0.506 (0.582)	-0.795 (0.876)	0.637 (0.785)
Avg. Employment Year 2 before Welfare Entry	-0.029 (0.399)	-0.506 (0.710)	-0.972** (0.484)	-0.034 (0.384)	-0.513 (0.702)	-0.972** (0.477)	-0.016 (0.383)	-0.500 (0.703)	-0.858* (0.479)
Avg. Employment Year 3 before Welfare Entry	-0.384 (0.255)	-0.567** (0.253)	-0.610** (0.242)	-0.419 (0.255)	-0.595** (0.252)	-0.658*** (0.242)	-0.405 (0.254)	-0.583** (0.250)	-0.635*** (0.238)
Prop. Never Employed Yrs 1-3 before Welfare Entry	-1.220 (0.952)	-4.269*** (1.653)	-5.221*** (1.859)	-0.533 (0.961)	-4.112** (1.640)	-4.710*** (1.807)	-0.550 (0.962)	-4.124** (1.641)	-4.922*** (1.833)
Avg. Earnings Year 1 before Welfare Entry/\$1,000	0.072 (0.137)	0.286 (0.188)	-0.057 (0.214)	0.061 (0.136)	0.284 (0.184)	-0.063 (0.210)	0.065 (0.135)	0.287 (0.183)	-0.026 (0.200)
Avg. Earnings Year 2 before Welfare Entry/\$1,000	-0.006 (0.106)	0.043 (0.185)	0.177 (0.134)	0.006 (0.103)	0.050 (0.182)	0.183 (0.135)	0.002 (0.103)	0.047 (0.183)	0.160 (0.136)
Avg. Earnings Year 3 before Welfare Entry/\$1,000	0.095 (0.061)	0.129** (0.050)	0.127*** (0.044)	0.100* (0.060)	0.136*** (0.050)	0.132*** (0.043)	0.097 (0.059)	0.133*** (0.050)	0.127*** (0.043)
Dummy CA TANF (CalWORKs) Period				0.467*** (0.169)	0.131 (0.174)	0.419*** (0.161)	0.465*** (0.169)	0.129 (0.174)	0.430*** (0.163)
Constant	-16.911*** (4.829)	-10.843* (6.469)	-1.680 (8.932)	-22.013*** (5.483)	-10.746 (7.134)	-6.864 (9.256)	-22.026*** (5.546)	-10.782 (7.141)	-7.689 (9.228)
Observations	548	548	548	548	548	548	548	548	548
Pseudo R-squared		0.137			0.137			0.137	
County Dummies	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Fiscal Year Dummies	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Robust standard errors between parentheses (adjusted by county/fiscal year clusters)

* significant at 10%; ** significant at 5%; *** significant at 1%