

NBER WORKING PAPER SERIES

PROGRAM EVALUATION AS
A DECISION PROBLEM

Rajeev H. Dehejia

Working Paper 6954
<http://www.nber.org/papers/w6954>

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

I am grateful to Gary Chamberlain, Edward Glaeser, Caroline Minter Hoxby, Guido Imbens, and Lawrence Katz for their support and encouragement in developing this project; to Joshua Angrist, Richard Blundell, and Jeffrey Smith for detailed comments and suggestions; to Gordon Anderson, Vivek Dehejia, Roberta Gatti, Kei Hirano, Jeffrey Liebman, Emily Mechner, Carl Morris, Donald Rubin, and Amartya Sen for their invaluable input; and to seminar participants at Columbia University, the Canadian Econometrics Study Group, the European Econometric Society Meetings, Harvard University, Johns Hopkins University, Mathematica Policy Research, McGill University, the NSF-NBER Bayesian Meetings, Ohio State University, Penn State University, Universitat Pompeu Fabra, University College London, and the University of Toronto for comments. I owe a special debt to Sadek Wahba for bringing the GAIN data to my attention and for many conversations during our ongoing collaboration, and to the Manpower Development Research Corporation (MDRC) for facilitating and permitting the use of these data. Responsibility for any remaining errors and omissions is my own. The views expressed here are those of the author and do not reflect those of the National Bureau of Economic Research.

© 1999 by Rajeev H. Dehejia. All rights reserved. Short sections of text, not to exceed two paragraphs, may be quoted without explicit permission provided that full credit, including © notice, is given to the source.

Program Evaluation as a Decision Problem

Rajeev H. Dehejia

NBER Working Paper No. 6954

February 1999

JEL No. C11, I38, J31

ABSTRACT

I argue for thinking of program evaluation as a decision problem. In the context of California's GAIN experiment (a randomized trial of a welfare-to-work alternative to AFDC), I show that GAIN first-order stochastically dominates AFDC when considering the choice between the treatment and control programs in terms of average earnings, even though the treatment effect is not statistically significant. I also argue for incorporating the post-evaluation assignment mechanism for the program under consideration into the evaluation process. I show that if policies, such as allowing a career counselor to choose which program individuals join, are included in the evaluation, then GAIN is superior to AFDC, whereas the opposite ranking emerges from the standard treatment versus control comparison which ignores potential heterogeneity in the treatment impact.

Rajeev H. Dehejia

Department of Economics

Columbia University

420 W. 118th Street, Room 1022

New York NY 10027

and NBER

dehejia@columbia.edu

“Clinton Signs Bill to Cut Welfare and Change State Role”

The New York Times, August 23, 1996

In a sweeping reversal of federal policy, President Clinton on Thursday ended six decades of guaranteed help to the nation's poorest children by signing into law a vast new welfare overhaul requiring the 50 states to deal more directly with the social burdens and the budget expense of poverty.

"Today we are taking a historic chance to make welfare what it was meant to be: a second chance, not a way of life," Clinton declared in signing the measure, which will affect tens of millions of poor Americans, largely by mandating work requirements and imposing a five-year lifetime limit on welfare help to needy families....

The heart of the complex new law abolishes Aid to Families With Dependent Children, the government's welfare bulwark, which provides monthly cash benefits to 12.8 million people, including more than 8 million children.

This is to be replaced by a system of block grants and vast new authority for the states, in the hope that they can fashion new work and welfare programs to solve the long-intractable problem of dependence on government.

Job creation will be a particular state burden, since the law requires most poor adults to find a job within two years of first receiving aid.

1. Introduction

This paper evaluates the Alameda portion of the Greater Avenues for Independence (GAIN) program with the aims both of understanding the impact of the program and of offering new methodological perspectives on program evaluation. An understanding of GAIN is valuable not only because it is very similar to California's current welfare program (CalWORKs), but also because similar welfare-to-work programs have been initiated by many states since the 1980s (Greenberg and Wiseman [1992] survey 24 such programs). At another level, GAIN is but one in a long line of social experiments (see Burtless [1995] for a recent survey), and methodological conclusions about evaluating this program will be of broad relevance.

Program evaluation typically is carried out by comparing the values of a range of outcomes of interest between the treatment and control groups, and considering the statistical significance of these treatment impacts.^{1,2} For example, compared with AFDC, the GAIN program³ causes a positive but not statistically significant increase in earnings and the probability of employment, and a significant increase in government welfare expenditure. Usually such differences are considered for various subsets of the sample.

The methodology I adopt differs because it models program evaluation as a decision problem. I focus on two aspects of what is in reality a complicated decision. First, in

¹ There are many relevant issues other than obtaining unbiased estimates of the treatment effect, such as the interaction of local conditions and implementation-related issues on the outcome. See Heckman (1992) and Heckman and Smith (1995).

² It is well known that in randomized trials such comparisons give unbiased estimates of the treatment effect (see Fisher [1935] and Neyman [1935]). The debate on the non-experimental evaluation side is more controversial; see Dehejia and Wahba [1996], Heckman [1989, 1990], Heckman and Robb [1985, 1986], Lalonde [1986], Manski [1989, 1993], and Manski and Garfinkel [1992], *inter alia*.

³ I will use the term GAIN to denote the GAIN package of services (including AFDC), and AFDC to denote the basic AFDC package.

choosing between the treatment and control programs, I pay particular attention to uncertainty about the outcome of interest, and how that uncertainty affects the choice between programs. It is well-recognized that a t-statistic does not embody all the information relevant for a rational decisionmaker.⁴ I instead use Bayesian predictive distributions – distributions which capture all of the uncertainty about the outcome of interest – which then allow for the use of standard expected utility theory in comparing the distribution of outcomes under treatment and control. Second, I consider how the programs being evaluated will be made available in a post-evaluation context. For example, will all individuals be required to participate in either the treatment or the control? Or is there scope to assign participation on an individual basis? I incorporate a range of post-evaluation assignment decisions into the process of evaluating the program by considering the (hypothetical) decision of a career counselor in assigning individuals between programs.

Such issues have been largely ignored in the evaluation literature. Two important exceptions are Heckman and Smith (1998), which offers a detailed discussion of social welfare criteria and the assumptions needed to identify them in the data (see also Heckman, Smith, Clements [1997]), and Manski (1995), which considers non-parametric bounds on the treatment impact when the treatment is non-homogeneous and post-evaluation assignment varies. The approach in this paper is complementary because I adopt a parametric approach which then allows me to focus on the individual-level decision problem and incorporating uncertainty into it. Of course, the notion that there is het-

⁴ The finance literature has made a similar point in a very different context. See Kandel and Stambaugh (1996). In addressing the questions, “Are stock market returns predictable and does it matter?” they argue that rather than formulating the question in terms of the statistical significance of the relevant parameters in an econometric model, one should look at the impact of such predictability on the portfolio decision of

erogeneity in the treatment impact is not original to this paper. The contribution of this paper lies in analyzing this heterogeneity through the individual-level decision.

Using the GAIN data, I demonstrate that the two methodological contributions outlined above are important in understanding the impact of the GAIN treatment. I show first that a career counselor who maximizes participants' post-treatment probability of employment will assign less than half of the individuals into GAIN. In terms of the evaluation, this implies that the policy of assignment by a career counselor yields higher average post-treatment earnings than either of the two policies (assigning all individuals into either GAIN or AFDC) that are normally considered. When it is selectively available through a career counselor GAIN emerges as viable in a cost-benefit sense as well, which overturns the traditional evaluation of this program. More generally, whenever there is heterogeneity in the treatment impact, such policies will be of central interest.

Second, I demonstrate that the evaluation of the GAIN program changes significantly when one accounts for uncertainty in a consistent manner. In particular, I show that the ranking that emerges between policies -- for example, assignment by a career counselor dominates GAIN, which in turn dominates AFDC in terms of post-treatment earnings -- is economically significant in the sense that the predictive distribution of earnings under one program first-order stochastically dominates the other. Any risk-averse agent would make an unequivocal ranking. In contrast, the ranking that emerges from a more standard t-test on the difference in means is equivocal; the difference is mildly significant.⁵

interest. See also Barberis (1996), Chamberlain and Imbens (1996), Geweke and Keane (1996), and Rossi, McCulloch, and Allenby (1995).

⁵ A highly relevant issue which I do not discuss here is: to what extent can one extrapolate the result to other populations of interest and to other time periods? When treatment effects are estimated at the individual level, one can, in principle, extrapolate to other populations to the extent that they have the same

The paper is organized as follows. Section 2 briefly describes the GAIN program and experiment. Section 3 presents a snapshot of the findings from GAIN. Section 4 describes the econometric model that I use. Section 5 examines the decision problems faced by typical individuals. Section 6 discusses the social decision problem and the choice of social welfare functions. Section 7 examines the results of the model at the social level, and Section 8 concludes the paper.

2. The GAIN Program and the GAIN Experiment

The GAIN program began operating in California in 1986, with the aim of “increasing employment and fostering self-sufficiency” among AFDC recipients (see Riccio, *et al.* [1994]). In 1988, six counties -- Alameda, Butte, Los Angeles, Riverside, San Diego, and Tulare -- were chosen for an experimental evaluation of the benefits of GAIN. A subset of AFDC recipients (single parents with children aged six or older and unemployed heads of two-parent households) were required to participate in the GAIN experiment (see Table 1). For its evaluation Alameda further confined itself to long-term welfare recipients (individuals already having received welfare for two years or more).

Potential participants from the mandatory group were referred to a GAIN orientation session when they visited an Income Maintenance office (either to sign up for welfare or to qualify for continued benefits).⁶ As a result, the chronology of the data and subsequent results is in experimental time, rather than calendar time. No sanctions were used if

support in the space of pre-treatment variables as the original sample (assuming ignorable assignment). If the model is suitably specified, one can also extrapolate through time. These issues will be the subject of future research.

⁶ In some counties AFDC recipients were allowed to volunteer for the GAIN program, but these units are not included in the public use sample.

individuals failed to attend the orientation sessions. However, once individuals started in the GAIN program, sanctions were used to ensure their ongoing participation. At the time of enrollment into the program, a variety of background characteristics were recorded for both treatment and control units, including: demographic characteristics; results of a reading and mathematics proficiency test; and data on 10 quarters of pre-treatment earnings, AFDC, and food stamp receipts.⁷

Of those who attended the orientation session, a fraction were randomly assigned to the GAIN program,⁸ and the others were prohibited from participating in GAIN.⁹ Each of the counties randomized a different proportion of its participants into treatment, ranging from a 50-50 split in Alameda to an 85-15 split in San Diego (see Table 1). Because assignment to treatment was random, the distribution of pre-treatment covariates is balanced across the treatment and control groups; the data legend lists each of the covariates.¹⁰ In terms of the chronology of data gathering, “experimental” time (which I also re-

⁷ Data on AFDC and Food Stamp receipts were taken from each county’s welfare records. Data on earnings were taken from the California State Unemployment Insurance Earnings and Benefits Records. Other background characteristics were taken from California’s client information (“GAIN-26”) form. See Riccio, *et al.* (1994).

⁸ The randomization was (as far as we know) independent of pre-treatment covariates. A different fraction was randomized into treatment in each county. See Table 2.

⁹ Of course, these individuals could participate in non-GAIN employment-creating activities. The existence of non-GAIN activities is important in interpreting the treatment effect from GAIN. The treatment effect measures the increase in earnings, employment, etc., from the availability of and encouragement (or requirement) to use GAIN-related activities compared with pre-existing employment services. To the extent that AFDC benefits are received by both groups, the real comparison at one level is between the two differing packages of supplementary services and requirements.

¹⁰ Since participants were randomized after registration, we expect to find that the distribution of background characteristics is (up to sampling variation) the same across the treatment and control groups within counties; this is partly true. Let b be a vector of pre-treatment characteristics. Assuming that $\bar{b}_T \stackrel{A}{\sim} N(\mathbf{m}_T, \frac{1}{n_T} \Sigma)$ and $\bar{b}_C \stackrel{A}{\sim} N(\mathbf{m}_C, \frac{1}{n_C} \Sigma)$, where T and C represent the treatment and control groups, n_T and n_C represent the number of treated and control units, and the \mathbf{m} ’s are k -vectors, we have

$(\bar{b}_T - \bar{b}_C) \left(\left(\frac{1}{n_T} + \frac{1}{n_C} \right) \hat{\Sigma} \right)^{-1} (\bar{b}_T - \bar{b}_C) \stackrel{A}{\sim} \mathbf{c}^2(k)$. The p-values are: Alameda 0.3364, Butte 0.5346, Los Angeles 0, Riverside 0, San Diego 0, and Tulare 0.0784. There are many possible reasons for the seeming lack of

fer to as “post-experimental” or “post-treatment” time) begins when individuals attend the GAIN orientation session. The early stages of experimental time thus coincide with the education and training part of the GAIN program.¹¹

In the GAIN experiment, the treatment is participating in the GAIN program; the control is receiving standard AFDC benefits. The GAIN program works as follows: based on test results and an interview with a case manager, participants are assigned to one of two activities. Those deemed not to be in need of basic education are referred to a job search activity (which lasts about three weeks); those who do not find work are placed in job training (which includes vocational or on-the-job training, and paid or unpaid work experience, lasting about three to four months). Those deemed to be in need of basic education may choose to begin job search immediately, but if they fail to find a job they must register for preparation toward either the General Educational Development certificate, Adult Basic Education, or English as a Second Language programs (lasting three to four months).¹² Participants were exempted from the requirement to participate in GAIN activities if they found work on their own.¹³

randomization in the three larger counties, including the possibility that randomization proportions may have varied across administrative offices within each county.

¹¹ More precisely, individuals were registered in the first quarter of experimental time. This means that in some cases the first quarter of experimental time in fact includes information from one or two months prior to the commencement of the experiment. For example, for an individual who attended an orientation session in February 1989, the first quarter of experimental time is from January to March 1989. Of course, some part of the first and second quarters could be spent participating in treatment activities. Pre-treatment data would cover the 10 quarters from July 1986 to December 1988.

¹² The public use data do not contain information on each individual’s participation in the various components of the program.

¹³ Note that only about 85 percent of the treated units actively participated in any GAIN activities (though by virtue of being in the GAIN sample they did attend an orientation meeting); the balance satisfied the requirements of the GAIN program on their own (in most cases finding employment within the first two or three quarters of experimental time). Thus, as observed earlier, this is important in interpreting the treatment effect as a comparison between earnings, employment, etc., when individuals are required to find a job or to participate in GAIN-related activities, and when they are not obliged to find jobs and only pre-existing employment-related services are available.

3. The GAIN Data in Alameda

From the six counties available in the GAIN experiment, this paper focuses on Alameda county. I briefly review the pre-treatment characteristics of the sample, and analyze the treatment effect on earnings and employment status. Finally, I touch on some examples of heterogeneity in the treatment effect.

Table 2 presents a profile of the participants in Alameda county: 85 percent are women, who on average have more than two children; the mean level of education is grade 10; and a quarter have previously participated in training programs. The average level of pre-treatment earnings is very low, ranging from \$150 to \$190 per quarter, but because 87 percent of pre-treatment earnings are zero, the average of non-zero pre-treatment earnings is higher, on the order of \$1,110 per quarter.¹⁴

Table 3 explicitly shows the impact of GAIN on the probability of unemployment for the 13 post-treatment quarters. In the first post-treatment quarter GAIN's impact is negative; this is not surprising since treatment units are participating in training activities in the first quarter. After a few quarters the treatment effect increases to the range of 4 percent to 6 percent (and is statistically significant).¹⁵

Figure 1 displays the average treatment effect on earnings for 13 quarters for each of the six counties. The figures illustrate the basic conclusion that GAIN participants in most counties enjoyed higher post-treatment earnings than their control counterparts. Ta-

¹⁴ Seven individuals are excluded from the original sample because of apparent coding errors in their covariates. These seven individuals are either coded as having 70 children or a previous hourly wage of more than \$300.

ble 4 explores the impact for Alameda in greater detail using an OLS regression of earnings on treatment status and covariates. We see a pattern similar to that found in Table 3: the treatment effect starts out negative, and then increases in most of the subsequent periods (with a number of statistically significant effects).

The assumption of a constant treatment effect across all individuals is very restrictive and unrealistic. The average treatment effects considered in the previous section potentially embody an array of heterogeneous treatment effects. Two examples illustrate this point. Figure 2 explores the interaction between the treatment effect and the score on the reading test: individuals who score 200 or more enjoy a higher treatment effect (although the standard error is quite large). In Figure 3, we see that individuals who have previously participated in job training programs also enjoy a higher treatment effect. Many such interactions potentially exist in the data. Thus, in subsequent sections the econometric model will allow for a heterogeneous treatment effect.

4. A Model of the Data

4.1 The Statistical Model

In my implementation I model earnings, given that AFDC and food stamps in principle are determined by non-stochastic rules. I denote the earnings component of the vector of outcome variables as $Y_{i,j}^t$, where $j=1$ (GAIN) or 0 (AFDC). $Y_{i,1}^t$ is interpreted as individual i 's earnings in period t if she was in GAIN, and $Y_{i,0}^t$ as her earnings if she was in AFDC; obviously one of these is counter-factual. Thus, observed earnings are defined as:

¹⁵ An earlier version of this paper (Dehejia [1997]) examines the impact of lagged employment status on the treatment impact, and shows that the treatment increases the probability of transition from unemploy-

$$Y_{it} = T_i Y_{i1}^t + (1 - T_i) Y_{i0}^t,$$

where T_i is a treatment indicator (=1 if individual i was assigned to GAIN, and =0 if she was assigned to AFDC). Realizations of the random variable are denoted in lower case, y_{it} .

A key feature of the distribution of earnings, which influences the model choice and was highlighted in Section 3, is the mass point in the distribution of earnings at zero. The strategy adopted is to model the probability of positive earnings and the distribution of positive earnings separately. For the former, a probit model is used for an indicator for positive earnings; for the latter, log earnings are modeled using a mixture of normals. Since individuals were randomly assigned into treatment and control, the model does not have to correct for sample selection bias in assignment to treatment.

Define an indicator variable, y_{it}^* , for positive earnings:

$$y_{it}^* = \begin{cases} 1 & \text{if } y_{it} > 0 \\ 0 & \text{if } y_{it} = 0, \end{cases}$$

and consider a latent variable, Y_{it}^{**} , which determines which value Y_{it}^* takes on:

$$y_{it}^* = \begin{cases} 1 & \text{if } Y_{it}^{**} > 0 \\ 0 & \text{if } Y_{it}^{**} < 0. \end{cases} \quad (1)$$

Assume

$$Y_{it}^{**} \left| \left\{ Y_{it}^* = y_{it}^* \right\}_{t=1}^{t-1}, \left\{ X_{it} = x_{it} \right\}_{t=1}^t, \mathbf{b} \sim N(x_{it} \mathbf{b}, 1), \quad (P)$$

for $i=1, \dots, I$ and $t=1, \dots, t$ (the probit model). The vector of explanatory variables is given by $x_{it} = \{1_{1it}, \dots, 1_{\tau it}, [1_{1it} \dots 1_{\tau it}] \cdot T_i, Z_i \cdot T_i, 1(Y_{i,t-1}=0)\}$, a $(2t + k + 1)$ -vector of regressor variables. $[1_{1it} \dots 1_{\tau it}]$ is a set of indicator variables for each quarter of post-experimental time

ment to employment. For a more detailed analysis of this issue, see Ashenfelter and Card (1985).

($1_{jt}=1$ if $t=j$, =0 otherwise), giving each period its own intercept. The treatment indicator is interacted with $[1_{1it} \cdots 1_{tit}]$. Since each period corresponds to experimental, rather than calendar, time, the treatment dummies produce a profile of the treatment effects over 13 quarters. Interactions between k exogenous regressors, Z_i , and the treatment indicator are also included, which allow the treatment effect to vary with observable pre-treatment characteristics. These characteristics include: indicators for the age and number of children, ethnicity, educational attainment, score on the reading and mathematics tests, sex, an indicator for previous participation in other training programs, 10 periods of pre-treatment earnings history, and a calendar time trend. The model allows for persistence in a very simple form, through lagged earnings, in particular an indicator for zero lagged earnings.

For positive earnings, a mixture-of-normals likelihood is adopted for log earnings:

$$\begin{aligned} \tilde{Y}_{it} \left| \left\{ Y_{it}^* = y_{it}^* \right\}_{t=1}^{t-1}, \left\{ X_{it} = x_{it} \right\}_{t=1}^t, \mathbf{g}, \mathbf{q}, \mathbf{s}, t, p \right. \\ \sim pN(x_{it}\mathbf{g}, \mathbf{s}^2) + (1-p)N(x_{it}\mathbf{g} + \mathbf{q}, \mathbf{t}^2), \end{aligned} \quad (\text{M})$$

where $\{\tilde{Y}_{it}\}$ is the log of the positive elements of $\{Y_{it}\}$. Note that (M) conditions only on the indicator for positive lagged earnings (rather than the level of lagged earnings).

I use diffuse priors for the parameters of both models. These are discussed in Appendix A, which also discusses the estimation procedure in detail.

4.2 The Predictive Distribution

Because the decision problems associated with program evaluation are in the space of outcomes, not the space of the parameters of the model, it is important to construct a distribution which embodies all of the uncertainty from the model and its estimation in the out-

come space. The estimation procedure, a Gibbs sampler, produces draws from the posterior distribution of the parameters of the model. For each model, we can use these draws to simulate the (posterior) predictive distribution of the outcomes; i.e., from (P) a distribution for the probability of positive earnings and from (M) a distribution for the level of positive earnings, in which we integrate out for the uncertain parameters.

Conditional on parameters, we can simulate the distribution by drawing for $Y_{I+1,1}$ (from the likelihoods (P) and (M)) and substituting forward for earnings in subsequent periods. To obtain the predictive distribution, we must account for parameter uncertainty; thus, we use draws from the posterior distribution of Θ (obtained from the two Gibbs samplers outlined in the appendix). For each draw from $p(\Theta|Data)$, we simulate the distribution from the likelihoods.

Thus, using this procedure, we obtain the joint predictive distribution of earnings for individual $I+1$ from periods 1,...,13. Given the values of the covariates, this predictive distribution represents the distribution of possible values for earnings incorporating all uncertainty (see the Appendix for further details).

4.3 The Choice and Fit of the Model

One issue in model choice, given the use of probit and mixture-of-normals likelihoods, is the use of the indicator for positive lagged earnings, rather than simply lagged earnings. For the GAIN data, a standard AR(1) does not capture the dynamics of the earnings process well. In particular, for individuals with positive earnings in a given period, in the following period it tends to over-estimate the persistence in their level of earnings. When extrapolating forward 10 or 12 periods, the process dramatically overstates the level of

earnings. Hence, I use an alternative specification, with an indicator for when lagged earnings are positive.

Another major issue is the choice of likelihood. Figures 4 and 5 give a sense of the fit of the model. These figures show the density of the empirical distribution of earnings for treated and control units (estimated through a histogram), averaged over the 13 post-treatment quarters, and plot the density of the predictive distribution of earnings. The joint predictive distribution of earnings across 13 periods is averaged over the individuals in the Alameda sample and over the 13 periods to produce the average predictive distribution, which is then plotted.

The use of a probit to model the probability of positive earnings is not controversial (other choices such as a logit yield similar results, and the probit is computationally convenient). As we see from the figures, the model fits the mass point at zero with substantial accuracy. For positive earnings, the figures illustrate the value of using a mixture of normals. The empirical distributions of log earnings for treated and control units are well approximated by the mixture because, for both distributions, the mixture is able to reflect the skewness of the empirical distribution. In fact, with a single normal, the predictive distribution of earnings would have a much thicker upper tail than the empirical distribution.^{16,17}

¹⁶ Given the predictive distribution it is possible to test for the fit of various features of the distribution. These tests essentially reconfirm the impression conveyed by the figures. The predictive distribution has a greater mass at lower levels of earnings than the empirical distribution. (The hypothesis that the median of the empirical distribution corresponds to the median of the predictive distribution is rejected). However, the predictive distribution fits the spread of the empirical distribution very well. The discrepancy in the 95-5-percentile spread is 2 per cent and in the 75-25 spread is 3 per cent. A similar conclusion is reached for the control distribution. An additional mixture component might help to better fit the peak in the empirical distribution, but when combined with the probit model, the effect on the overall predictive distribution of earnings is minimal.

5. The Individual-Level Impact

This section studies the individual-level impact of the GAIN treatment. This not only provides a detailed view of the impact of the program, but also lays the foundations for the analysis of the social welfare problem in Section 7.

Imagine that a career counselor has to choose whether to assign an individual into GAIN and AFDC. For the career counselor, an individual is identified by her pre-treatment characteristics. Thus, the key assumption is that the individuals under consideration are exchangeable with those in the data, conditional on observable covariates: earnings for individuals with the same covariates are taken to be drawn from the same distribution. Given that the decision is being made by a career counselor, not the individual herself, the assumption seems reasonable.¹⁸

5.1 Two Typical Examples

Tables 5a and 5b list the pre-treatment covariates of two individuals from the Alameda county sample for whom we see typical patterns in the distributions of earnings under treatment and control (one being a clear winner from GAIN and the other a clear loser).

The first is a woman (“Ms. Thirteen Fifty-Three”), aged 42. She is the head of a single-

¹⁷ The choice of how many mixture components to use is heuristic: the use of more than two mixture components does not seem to improve the fit of the model. An alternative would be to use a method which endogenously selects the number of mixture components, but the added complexity is probably not warranted given that the empirical distributions are well behaved.

¹⁸ The assumption of exchangeability conditional on covariates is not unique to my application. This assumption, or some alternative, is needed any time we want to extrapolate from a dataset to a new situation. Unless information on important unobserved variables such as preferences is explicitly solicited, one must use the exchangeability assumption or some other equally strong assumption specifying the form of the lack of exchangeability. If the individual herself is making the choice, we must assume that any private information she has is independent of the observed covariates.

parent household, has one child between the age of 12 and 18, and has completed high school and three years of additional training. Her earnings history shows that she was employed in two of 10 quarters prior to the experiment. The second individual (“Ms. One”) is a 32-year-old woman, the head of a single-parent household, has one child between the age of 6 and 11 and two children between 12 and 18, and has an educational attainment of grade nine. Her earnings history shows zero earnings in each of the 10 pre-treatment periods. I consider each individual in turn.

Table 6a shows the probability of positive earnings and the mean and standard deviation of the predictive distribution of earnings (including the mass point at zero earnings) for each period under both treatment and control for Ms. Thirteen Fifty-Three. Her mean earnings in the first three periods are lower under treatment than under the control, but her probability of positive earnings in these periods is higher under treatment. Intuitively, even though she is more likely to find a job under the treatment program, any job she finds is not likely to pay more in the first three periods than a job she might have found otherwise. However, from the fourth period on, both earnings and the probability of employment are higher under treatment. The profile of the treatment effect is increasing, in a pattern similar to that depicted in Figure 1 for Alameda county on average. However, the standard deviation of control earnings is higher than that of treatment earnings, and the difference between the treatment and control earnings is small compared with the magnitude of the standard deviation.¹⁹ An important question is whether her risk attitude would affect her preference between the two programs.

¹⁹ The difference in means is not significant in the sense that the 95 percent probability intervals of the posterior distributions overlap. But the standard deviation of the predictive distribution is not very infor-

One way to address this would be to make an expected utility comparison from the predictive distributions of earnings under treatment and control. Figure 6 depicts the cumulative distribution functions for the predictive distribution for each of the 13 periods. These summarize all the information available for the individual to consider, by combining the predicted probability of positive earnings with the distribution of positive earnings. The figures are very revealing. In periods one to three, the control earnings (almost) first-order stochastically dominate the treatment earnings; and in periods three to 13, treatment earnings unambiguously dominate control earnings. This is a simple illustration of the fact that even when the means of the two distributions under consideration are not very different in terms of t-statistics, the underlying decision between the two may be clearcut. In this case, as long as the individual does not have an extremely high discount rate, we would advise her to join the treatment group.²⁰ Whether she, in fact, would follow this advice depends on the factors she considers, which may or may not be part of the model.

For the second individual, matters are different. We see in Figure 7 that her distribution of earnings under the control first-order stochastically dominates her distribution under treatment in each period. As long as she prefers more earnings to less, she unambiguously would be advised not to participate in GAIN.

Of course, first-order stochastic dominance does not suffice to compare all the distributions which arise. In general, expected utility comparisons would be required.

mative, because of the mass point in the distribution. This is another reason to examine the entire distribution of earnings, which we do below.

²⁰ What would we advise on whether she should go to school, train, etc.? Since our data contains no information regarding this issue, we assume that the choices she makes would not be too different from those of similar people represented in the sample.

5.2 The Importance of Accounting for Uncertainty

A natural question which arises from the preceding analysis is: would similar decisions have been reached if uncertainty had not been accounted for as comprehensively? In particular, one might imagine using the model described in Section 4, but, rather than using the full posterior distribution of the parameters, using point estimates and treating them as though they were the true parameters. Of course, even ignoring parameter uncertainty, the intrinsic uncertainty embodied in the likelihoods of (P) and (M) has to be taken into account. Table 6b considers such an exercise for Ms. Ten, whose characteristics are given in Table 5c. Columns 1 to 4 of Table 6b present the distribution of her earnings in each of the 13 quarters, ignoring parameter uncertainty, but still accounting for the uncertainty conditional on parameters. In contrast, columns 5 to 8 present the posterior distribution of her earnings, in which parameter uncertainty is accounted for. The means of the two sets of predictions are broadly similar, as are the standard deviations. Of course, since the underlying distributions are highly non-normal, the first two moments are only partially informative. The two distributions can be compared directly through their expected utilities. An expected utility comparison for log utility and constant relative risk aversion (CRRA) preferences (with coefficient of relative risk aversion equal to 3) reveals that the advice to Ms. Ten differs depending on which distribution is used. Ignoring parameter uncertainty, her expected utility is higher under GAIN; instead her expected utility under the predictive distribution is higher under AFDC.

Of course this example was chosen precisely because ignoring uncertainty leads to different advice than accounting for it. In cases where the two distributions are starkly different, ignoring uncertainty would not, typically, lead to a change in decision. For the

overall sample from Alameda, uncertainty affects the decisions of about 10 percent of individuals.

5.4 Heterogeneity and the Value of Choice

We could consider such decision problems for a wider array of individuals. The differences in the results would reflect the underlying heterogeneity in the treatment effect. One view of this is presented in Table 7. Assume that each of the 1,360 individuals in the Alameda sample is offered the choice between GAIN and AFDC, and, as in the previous examples, imagine that these individuals are advised on the choice based on the predictive distributions of their earnings in each period under each program.²¹ For the moment, I suppose that the advice is based on (time-additive) constant relative risk aversion preferences (with a coefficient of relative risk aversion equal to 3).

Table 7 presents the mean of expected post-treatment earnings under GAIN and AFDC and the mean of pre-treatment covariates for two sub-groups of the sample: the group that “chooses” (i.e., is advised to enter) GAIN over AFDC (for whom expected utility from post-treatment income is higher under GAIN than under AFDC), and the complement of this group.²² Only 40 percent of the sample fare better (in an expected utility sense) under GAIN than AFDC. The comparison of these two groups is revealing. Those benefiting from GAIN have fewer children on average (except between the age of 6

²¹ As mentioned in the previous section, I model only earnings. When needed, I use a deterministic rule to fill in suitable amounts for AFDC and food stamps. For AFDC, only six individuals in the Alameda sample have both zero income and zero AFDC benefits. About 10 percent of the sample have both positive earnings and positive AFDC benefits. The balance, about 90 percent of the sample, are receiving full AFDC benefits, since they have zero earnings. Because all of the data needed to determine AFDC benefits are not available in the dataset, I use a fixed-effects regression model to predict AFDC and food stamp benefits for these individuals.

and 11), have higher scores on the reading and mathematics tests, have a higher level of educational attainment, are five years younger, and half of them have previously participated in training programs. Of particular note is the difference in the level of pre-treatment earnings. Individuals who are advised to enter GAIN have average quarterly earnings on the order of \$300, while those who advised not to enter GAIN have quarterly earnings on the order of \$50.

Table 7 offers one view of the value of choice to individuals in the sample. For individuals with high pre-treatment earnings (the primary beneficiaries of GAIN), the potential value is high: their expected post-treatment earnings under GAIN are more than \$300 per quarter higher than under AFDC. In contrast, the potential value of choice of programs is much lower for individuals with low pre-treatment earnings: the difference between their expected quarterly earnings in the two programs is \$20 per quarter.

6. The Social Choice Problem

Thus far the analysis has focused on the individual-level decision between GAIN and AFDC. This section takes the next step by asking how the policymaker can decide which program or combination of programs to make available, given the pattern of individual effects. There are two aspects to this decision. First is the set of policies under consideration, where policies are defined as rules for determining each individual's (post-evaluation) assignment to treatment. I consider four alternatives: (1) All individuals are required to participate in GAIN; (2) All individuals remain in AFDC; (3) A career counselor assigns each individual into the program which is most likely to give him or her posi-

²² The administrative cost of GAIN is estimated at \$3,638 for 13 quarters, based on chapter 3 of Riccio, *et*

tive earnings, where the decision is informed by the model in the previous sections (that is, a policy “mandated” to increase employment); and (4) Each individual is allowed to “choose” participation in AFDC or GAIN (i.e., individuals are assigned to GAIN or AFDC based on the expected utility of the distribution of their earnings under each program, where they are assumed to be either risk neutral or have CRRA preferences with the coefficient of relative risk aversion, q , equal to 1 or 3).²³

The second aspect of the social decision is the set of criteria (social welfare functions) that the policymaker uses to decide which policy to adopt. I confine the policymaker to concerns about average outcomes such as earnings and welfare costs and concerns about inequality in the distribution of earnings.²⁴ The first set of concerns is typically captured in the following social welfare functions (SWFs): (1) average (labor) earnings per individual; (2) the average (undiscounted) value of expected income (including labor earnings, AFDC, and food stamps) across individuals and time; (3) the fraction of individuals employed (having positive income) averaged across time periods; (4) the average expenditure per period per person entailed by a given program; and (5) the average earnings per individual minus total cost in excess of standard AFDC services (the difference in this criterion can be interpreted as the increase in earnings minus the increase in cost). The latter concerns are captured through three standard social welfare functions of the form:

$$\sum_i g\left(E\left(u\left(Y_{i T_i}\right)\right)\right),$$

al (1996).

²³ Caballero (1991) suggests that a reasonable value for the coefficient of relative risk aversion is between 2 and 3. He cites Hall (1988) and Friend and Blume (1975).

²⁴ In this analysis the policymaker is confined to SWFs that do not depend on the joint distribution of the outcomes because this cannot be identified without further information or stronger priors. See Heckman, Smith, and Clements (1997).

where the expectation is over the predictive distribution of earnings for a given set of post-treatment assignments $\{T_i\}$ and $g(\cdot)$ is specified according to: $g(x)=x$, utilitarian; $g(x)=\log(x)$, log; or $g(x) = (x^{1-\epsilon} / 1 - \epsilon)$, where $\epsilon=3$, an intermediate case (see Deaton and Muellbauer [1980]). I also consider a Rawlsian social welfare function of the form:

$$\min_i \left\{ E \left(u \left(Y_{i T_i}^t \right) \right) \right\}.$$

These social welfare functions are computed over the predictive distribution of earnings which yields predictive distributions of these social welfare functions, distributions which capture the full range of uncertainty regarding their outcomes.

7. What We Learn from GAIN

A useful benchmark for the social evaluation of GAIN is the conclusion one would reach using differences in means rather than the predictive distributions which I consider below. These are presented in Table 8 for the six social welfare criteria discussed in the previous section. For post-treatment labor earnings, total income per quarter, and the probability of employment, there is a positive but insignificant treatment impact. For total fiscal expenditure, earnings net of costs, and total expenditure net of tax receipts, there is a significant increase in costs.

Tables 9a, 9b, and 9c apply the social welfare analysis outlined in Section 7 to the predictive, rather than empirical, distributions of outcomes under treatment and control, allowing for a range of post-evaluation assignment mechanisms. I simulate the predictive distribution of the earnings under treatment and control for 13 quarters of post-treatment earnings for each of the 1,360 individuals in the sample, and where necessary use determi-

nistic rules to impute values for AFDC receipts and food stamps. Individuals are assigned to treatment and control groups as outlined in Section 6.

Table 9a uses SWFs (1) to (5) to assess the policies outlined in Section 7. Each SWF-policy combination defines a cell, within which the social welfare function is computed for the entire predictive distribution of earnings. Table 9a presents the mean and the 2.5 and 97.5 percentiles of this distribution for each cell.

Consider first SWF(1), average post-treatment earnings per person per quarter. From the first two cells of column 1, the mean predictions from the posterior of the model (\$530 for GAIN and \$405 for AFDC) are similar to those obtained from the empirical distribution, within \$80 for GAIN and within \$30 for AFDC. The 95 percent posterior confidence intervals of expected quarterly earnings under treatment and control just overlap. Cells 3 and 4 show that the policies of mandating individuals to enroll in either GAIN or AFDC based on their probability of post-treatment employment, or assigning them based on their expected utility of earnings (if they are risk neutral), both yield substantially higher average quarterly earnings than the policy of enrolling everyone in GAIN (\$624 and \$654 compared with \$530). At the individual level, however, the impacts of the policies are quite different. The mandated policy assigns 300 more individuals into treatment than the (risk neutral) “choice” policy. This difference arises because, for some individuals, while GAIN does increase the probability of employment, it does not increase average earnings; it steers some individuals toward jobs that pay less than those they might have found without having their job search or employment disrupted by training. When individuals exhibit a low degree of risk aversion ($q \approx 1$), the pattern of their choices is similar to

the risk neutral case. With $q=3$, more individuals prefer GAIN to AFDC. In both cases the average value of expected earnings is similar to the risk-neutral case.²⁵

In light of the analysis in Section 5, it is not surprising that the career-counselor mechanisms yields higher earnings than GAIN (or AFDC), because these mechanisms steer away many (about 700) individuals who are not expected to benefit from GAIN. The difference in earnings between the individual-level choice policies and the non-choice policies, \$50 to \$200 per person per quarter, provides one measure of the potential value to participants of being permitted to choose their program. The difference is large relative to average quarterly earnings (which are on the order of \$300).

The advantage of working with the entire distribution of social welfare values is seen in Figure 8, which is analogous to Figures 6 and 7. Comparing means and their 95 percent confidence intervals, one might conclude that no decisive statement could be made about the differences among policies considered in Table 9a: even though the means of the policies are different, their 95 percent confidence intervals overlap to varying degrees. But when we look at the entire distribution in Figure 8, we note that GAIN first-order stochastically dominates AFDC, and in turn is dominated by the choice and mandated policies. Thus, in comparing the entire distribution of values, a very strong ranking emerges based on SWF(1).²⁶

²⁵ The findings are much sharper than would be obtained using Manski's (1995) extreme bounds analysis of what he calls the "mixture problem." Of course, the sharper findings come at a price: the willingness to specify a likelihood model. But having paid the price the advantage is a full posterior distribution for the outcomes of interest, allowing for a richer analysis of individual decisions.

²⁶ It would not be a surprising conclusion that one distribution first-order stochastically dominates another given that their means differ. But the argument is that it is more informative to compare the entire predictive distributions of outcomes of programs than to compare the means of their empirical distributions, as is normally done.

In column 2, we see that adding AFDC receipts and food stamps to the analysis does not alter the ranking. GAIN produces “a larger pie” than AFDC (\$2640 compared with \$2506), with the mandated and choice mechanisms doing slightly better than GAIN (\$2728 and \$2756). When individuals are allowed to choose, we see a pattern similar to that in column 1.

Social welfare function (3) ranks the policy alternatives by the criterion of putting people to work. GAIN (in keeping with its stated mandate) does succeed in putting more people to work than AFDC, although the magnitude of the difference is not large (a 0.21, compared with 0.18, post-treatment probability of employment). The mandated policy succeeds in putting more people to work than GAIN (a 0.24 probability of employment) and slightly more than the choice (1) policy (which produces a 0.22 probability of employment). Again, when individuals are risk averse, we see a pattern similar to that in columns 1 and 2.

Social welfare function (4) considers the fiscal criterion, and reveals that GAIN is clearly the most expensive of the policies (\$5755 per person per quarter) and AFDC is the cheapest (\$2101); the posterior confidence intervals for these two policies do not overlap. To some extent this is not surprising, because GAIN differs from AFDC precisely in offering costly education and training, services that AFDC recipients would have to pay for at their own expense.²⁷ When participants are assigned individually, the pattern is similar to that found in column 1. Social welfare function (5) reveals that the increased earnings realized by assigning all individuals into GAIN do not offset the increased costs when compared with AFDC (a net difference of $-\$165$ per person per quarter), but do offset the

²⁷ Of course, this ignores the issue of non-GAIN sources of funding.

costs when a subset of individuals are assigned (or choose) to join GAIN (increased earnings net of costs are \$167). Figure 9 illustrates that the ranking of the first four policies implied by SWF(5) in fact amounts to first-order stochastic dominance.

Thus, combining SWF (1), (2), and (3) with SWF (4) and (5) suggests that the policies which allow choice regarding individual assignment to treatment and control dominate the policy in which they are all assigned to GAIN: such policies are cheaper and result in higher average earnings per person. Choice mechanisms also dominate AFDC, though in the more limited sense of increased earnings net of cost. With the caveat that we take individual utility to be a function only of earnings, we can conclude that allowing a career counselor to assign individuals into treatment and control groups makes both individuals and the social planner better off; thus we reach a positive assessment of the treatment. In contrast, ignoring the possibility of differential assignment, one would conclude that GAIN has a mixed and limited impact on individual earnings, with its benefits more than offset by the increased costs of the program.

Another set of concerns for the policymaker is the distribution of the benefits from GAIN (see Heckman, Smith, and Clements [1997] for a related analysis) . Table 9b examines the *ex post* inequality generated by the policies in Table 9a. It presents percentiles of the predictive distributions of earnings in each program (averaged over the 13 quarters and 1360 individuals). The 5th and 25th percentiles of the distribution for each of the five policies is zero. The differences among the policies emerge in the upper percentiles. The 75th percentile is lower for AFDC than for the choice policies (\$601 compared with \$772), but is higher than for GAIN (\$491). This underscores the point which arose in earlier discussion, namely that the benefits from GAIN are far from uniform. For many indi-

viduals, the expected earnings achieved through GAIN are lower than what they would have achieved through AFDC. The difference between GAIN and AFDC arises in the upper tail of the distribution, for example at the 95th percentile; this is depicted graphically in Figure 10. Looking at the difference between the 90th and 10th percentiles, we note that the voluntary policies produce more inequality than either GAIN or AFDC (a 90-10 spread of \$1616 compared with \$1252 and \$1312).²⁸

Table 9c examines *ex ante* inequality by applying a range of standard social welfare functions to the predictive distribution of post-treatment earnings. The utilitarian, Rawlsian, log, and intermediate social welfare functions discussed in Section 7.2 are applied to the certainty equivalents of the distributions of earnings for each individual. Thus, the social welfare rankings do not require any standard errors or confidence intervals; uncertainty is already accounted for in the certainty equivalents. Table 9c reveals that if the policymaker is sufficiently inequality-averse (as embodied in SWFs (7) to (9)), then he will prefer AFDC to GAIN. But, under both the inequality-neutral and inequality-averse specifications, the choice policies will be preferred to both GAIN and AFDC. This finding is not trivial. Although each individual must have weakly higher utility under a choice mechanism as compared to being assigned to GAIN or AFDC, depending on the policymaker's aversion to inequality, the social ranking may not mirror the individual ranking.

In summary, from Table 9a we learn that, the choice mechanism dominates both GAIN and AFDC in terms of earnings. In terms of welfare expenditures, AFDC is the cheapest policy, although the choice mechanism again dominates once the increased earnings realized through the treatment are netted out. In Table 9b we see that the benefits

²⁸ Because the joint distribution of earnings is not identified, we cannot make claims about how particular

from GAIN are enjoyed at the upper end of the earnings distribution, and that a choice mechanism produces more inequality in the distribution of earnings than either AFDC or GAIN. Finally, from Table 9c, we learn that even a policymaker who is inequality-averse would prefer the distribution under a choice mechanism to both GAIN and AFDC.

9. Conclusion

This paper examines the implications of shifting the emphasis in program evaluation from examining average treatment effects and their statistical significance to looking at the underlying decision problems. Two main differences emerge.

First, by looking at the entire distribution of earnings under treatment and control for a wide range of individuals, I conclude that for most individuals the choice between GAIN and AFDC is clear-cut, with over half of the sample preferring GAIN to AFDC. This contrasts with the conclusion one would reach by considering the statistical significance of differences in means, which are not significant for post-treatment earnings for the sample as a whole. The group benefiting from GAIN is identified as those individuals with higher pre-treatment earnings.

Second, the policy of assigning individuals into treatment and control based on pre-treatment characteristics dominates GAIN; it yields higher average earnings and leads to lower government expenditure. By the criterion of increased earnings net of increased costs, this policy also dominates AFDC. This illustrates the importance of thinking about how a program will (or can) be made available in a post-evaluation context, and of incorporating these possibilities into the evaluation of the program. If the possibility of assign-

individuals fare relative to the distribution in each program. See Heckman, Smith, and Clements (1997).

ing only some individuals into the treatment is ignored, then GAIN would be rejected by most of the social welfare criteria considered because the average increase in participants' earnings does not outweigh the increase in costs. But by allowing individuals to be differentially assigned into GAIN and AFDC, GAIN does substantially benefit some individuals; if the program were made available to them, it would be viable by most social welfare criteria.

These results are certainly important for an analysis of the GAIN data, but are also relevant to many other exercises in program evaluation. For any program in which there is heterogeneity in the treatment impact, there is potentially a role for differential assignment into treatment. This is especially true for programs in which the gains from the treatment do not exceed the cost for some individuals. Also, the importance of comparing outcomes under different programs using their predictive distributions rather than simply the first moments of their empirical distributions applies quite broadly to other evaluations. The relevance of this framework extends beyond the case of randomized experiments considered here. In non-experimental settings, if a sufficiently rich set of covariates were observed (justifying the assumption of selection on observable covariates) and if one conditions on them in a flexible way, then the same methodology can be adopted (see Rubin [1977, 1978]).

The model can be extended in a number of directions. First, in some policy contexts, there may exist substantial prior information regarding the control program. Such information could readily be incorporated into the priors of the model. Second, there is scope to add greater heterogeneity, perhaps by using a hierarchical model to incorporate many more interactions. Third, the model could be modified to forecast beyond the 13

quarters included in the dataset to extend the evaluation to longer horizons. Fourth, the individual “choice” framework of assignment by a career counselor could be extended to allow individuals to incorporate private information into their decisions; the policymaker then would not simply offer individuals a choice but would design incentive-compatible assignment mechanisms. These are subjects for future research.

Appendix A: The Priors and Estimation Procedure

A.1 The Estimation Procedure

The posterior distribution of the parameters of the two models (P) and (M) are obtained through Gibbs sampling procedures. The Gibbs sampler is a Markov chain Monte Carlo simulation technique which simulates the joint posterior of the parameters of the model. Instead of drawing directly from the joint posterior (often intractable), it draws successively from the posterior of each parameter (or block of parameters) conditional on all of the other parameters. For any starting values (given certain conditions), these draws will eventually converge to draws from the true posterior (see Geman and Geman [1984], Gelfand and Smith [1990], Tanner and Wong [1987], as well as Chamberlain and Imbens [1996], Gelman, Carlin, Stern, and Rubin [1996], and Rossi, McCulloch, and Allenby [1995]).

In many cases, such as the probit and mixture models, the task of drawing from the joint posterior is simplified by augmenting the parameter space of the model. For the probit model, the parameter space is expanded to include the latent variables y_{it}^{**} ; conditional on these, the probit model reduces to a standard regression model, and, conditional on all other parameters, it is easy to draw from the posterior distribution of y_{it}^{**} . For the mixture model, the parameter space is expanded to include indicators for which component of the mixture each observation is drawn from; again, conditional on these indicators, it is easy to update the other parameters, and *vice versa*.

A.2. The Probit Model

The Gibbs sampling algorithm for the probit model has been worked out by Albert and Chib [1993] (see also Chib [1992], and Chib and Greenberg [1994]).

We stack the observations in the form:

$$y_{it}^{**} = x_{it} \mathbf{b} + \mathbf{e}_{it}$$

for $y_{it}^{**} \in y_i^{**} = (y_{i1}^{**}, y_{i2}^{**}, \dots, y_{it}^{**})'$, for $i=1, \dots, I$, $I=1,360$, and $t=13$. Of course, y_{it}^{**} is not observed. The key to the Gibbs sampling procedure is that conditional on \mathbf{b} , it is easy to draw from the posterior distribution of y_{it}^{**} , and then using these to draw from the posterior of \mathbf{b} . Given diffuse priors for \mathbf{b} and an arbitrary starting value $\mathbf{b}^{(0)}$, the Gibbs sampling scheme is then:

(1) Conditional on $\mathbf{b}^{(j)}$, draw values for y_{it}^{**} : for $\{it : y_{it}^{**}=0\}$, from the negative portion of a normal distribution with mean $x_{it} \mathbf{b}^{(j)}$ and variance 1, and for $\{it : y_{it}^{**}=1\}$, from the positive portion of the same distribution. Denote the filled-in dependent variable $y_{it,Z}^{(j+1)}$, so that $y_{i,Z}^{(j+1)} = (y_{i1,Z}^{(j+1)}, \dots, y_{it,Z}^{(j+1)})'$.

(2) Conditional on $Y_Z^{(j+1)} = (Y_{1Z}^{(j+1)}, \dots, Y_{IZ}^{(j+1)})'$, draw for $\mathbf{b}^{(j+1)}$ from

$$N(\hat{\mathbf{b}}^{(j+1)}, (X'X)^{-1}),$$

where $\hat{\mathbf{b}}^{(j+1)} = (X'X)^{-1} X'Y_Z^{(j+1)}$, with $X_i = (x_{i1}, \dots, x_{it})'$, and $X = (X_1', \dots, X_I)'$.

From an arbitrary starting value, this is iterated 2,000 times, producing $(Y_Z^{(j)}, \mathbf{b}^{(j)})$. The first 500 iterations are discarded, leaving 1,500 draws from the posterior distribution of the parameters, which will be indexed $j=1, \dots, 1,500$.²⁹

A.3. The Mixture Model

When proper parametric priors are used, the Gibbs sampler for a mixture model is straightforward. It is, however, essential that priors be informative to some extent. Even though with parametric priors it is possible to identify the mixture, without sufficient prior information the computational algorithm can break down (see Robert [1996], and also Geweke and Keane [1996]). In setting up the Gibbs sampler, prior information is incorporated in two ways. First, the prior requires that $\mathbf{q} > 0$; a normal prior (with mean zero and large variance) is used, and truncated suitably. Second, prior information is provided about the variance of each component, \mathbf{s}^2 and \mathbf{t}^2 , in the form of the number of prior observations and the specified prior variance for each component. For \mathbf{s}^2 , the prior is 30 observations with a sample variance of 1.5, similar to the variance arising in a single component normal model for this data. The 5th and 95th percentiles of the prior are 1.03 and 2.41. Though this is reasonably tight, in the updating procedure the weight on the prior will be very low (see the discussion below). For \mathbf{t}^2 , the prior is 30 observations with a sample variance of 0.25, with the aim of picking up peaked segments observed in the empirical distribution. (The 5th and 95th percentiles of this prior are 0.17 and 0.41.)

A flat prior is used for \mathbf{g} . The prior on p is expressed in terms of the number of prior observations seen from each component of the mixture; these are set to 1 for each component. The priors can be summarized as: $p(\mathbf{g}) \propto c$, $\mathbf{s}^2 \sim \text{Inv-}\mathcal{C}^2(v_0, \mathbf{s}_0^2)^{30}$, $v_0 = 30$, $\mathbf{s}_0^2 = 1.5$, $\mathbf{t}^2 \sim \text{Inv-}\mathcal{C}^2(n_0, \mathbf{t}_0^2)$, $n_0 = 30$, $\mathbf{t}_0^2 = 0.25$, $\mathbf{q} \sim \mathcal{N}(\mathbf{q}_M, \mathbf{q}_V) \cdot I(\mathbf{q} > 0) / (1 - \Phi(-\mathbf{q}_M / \mathbf{q}_V^{1/2}))$, $\mathbf{q}_M = 0$, $\mathbf{q}_V = 100$, $p \sim \text{Beta}(\mathbf{a}_1, \mathbf{a}_2)$, $\mathbf{a}_1 = \mathbf{a}_2 = 1$.

The Gibbs chain generates a latent variable, z_i , which is an indicator for which of the two components a given observation originates from (1 if it is from the first component, 0 for the second). Conditional on z_i , the chain is straightforward, and conditional on all other parameters it is easy to update the z 's (see Robert [1996]). Take starting values $z^{(j)} = \{z_i^{(j)}\}$, $\mathbf{s}^{(j)}$, $\mathbf{t}^{(j)}$, $p^{(j)}$, $\mathbf{q}^{(j)}$, and $n_1 = \sum_i z_i^{(j)}$, $n_2 = \sum_i (1 - z_i^{(j)})$:

(1) Conditional on $z^{(j)}$, $\mathbf{s}^{(j)}$, $\mathbf{t}^{(j)}$, $p^{(j)}$, $\mathbf{q}^{(j)}$, draw for $\mathbf{g}^{(j+1)}$ from $N(\tilde{\mathbf{g}}_{GLS}, (\tilde{X}' \Lambda^{-1} \tilde{X})^{-1})$, where $\Lambda = \begin{bmatrix} (\mathbf{s}^2)^{(j)} I_{n_1} & 0 \\ 0 & (\mathbf{t}^2)^{(j)} I_{n_2} \end{bmatrix}$, $\tilde{\mathbf{g}}_{GLS} = (\tilde{X}' \Lambda^{-1} \tilde{X})^{-1} \tilde{X}' \Lambda^{-1} Y^{(j)}$, $Y^{(j)} = [Y_1^{(j)'} Y_2^{(j)'}]'$, $Y_1^{(j)} = \{\tilde{Y}_i : z_i^{(j)} = 1\}$, and $Y_2^{(j)} = \{\tilde{Y}_i - \mathbf{q}^{(j)} : z_i^{(j)} = 0\}$, with \tilde{Y}_i formed

²⁹ Several diagnostics suggest that throwing out the first 500 runs is sufficient to converge to draws from the posterior. These include considering a wide variety of starting points, running the sampler for more iterations, and comparing the mean of the posterior of the parameters with maximum likelihood estimates of the same parameters.

³⁰ If $\mathbf{s}^2 \sim \text{Inv-}\mathcal{C}^2(v_0, \mathbf{s}_0^2)$, then $(1/\mathbf{s}^2) \sim \mathcal{C}_{v_0}^2 / v_0 \mathbf{s}_0^2$.

from the log of the stacked, positive elements of y_{it} , and \tilde{X} formed from the elements of X corresponding to $Y^{(j)}$.

(2) Conditional on $z^{(j)}$, $\mathbf{g}^{(j+1)}$, $\mathbf{t}^{(j)}$, $p^{(j)}$, $\mathbf{q}^{(j)}$, draw for $(\mathbf{s}^2)^{(j+1)}$ from $\text{Inv} - \mathbf{c}^2\left(v_0 + n_1, \frac{v_0 \mathbf{s}_0^2 + n_1 s_1^2}{v_0 + n_1}\right)$, where $s_1^2 = \sum_{i=1}^{n_1} \left((Y_1^{(j)})_i - X_i^{(1)} \mathbf{g} \right)^2 / n_1$ and $X^{(1)}$ are the

elements of \tilde{X} corresponding to $z^{(j)}=1$. We see that the posterior weights the prior and the sample information by the number of prior and sample observations; with over 3,000 observations, clearly most of the weight will be on sample information.

(3) Conditional on $z^{(j)}$, $\mathbf{g}^{(j+1)}$, $\mathbf{s}^{(j+1)}$, $p^{(j)}$, $\mathbf{q}^{(j)}$, draw for $(\mathbf{t}^2)^{(j+1)}$ from $\text{Inv} - \mathbf{c}^2\left(n_0 + n_2, \frac{n_0 \mathbf{t}_0^2 + n_2 s_2^2}{n_0 + n_2}\right)$, where $s_2^2 = \sum_{i=1}^{n_2} \left((Y_2^{(j)})_i - X_i^{(2)} \mathbf{g}^{(j+1)} - \mathbf{q}^{(j)} \right)^2 / n_2$ and $X^{(2)}$

are the elements of \tilde{X} corresponding to $z^{(j)}=0$.

(4) Conditional on $z^{(j)}$, $\mathbf{g}^{(j+1)}$, $\mathbf{s}^{(j+1)}$, $\mathbf{t}^{(j+1)}$, $p^{(j)}$, draw for $\mathbf{q}^{(j+1)}$ from $N\left(\frac{\frac{1}{q_v} \mathbf{q}_M + \frac{n_2}{(\mathbf{t}^2)^{(j+1)}} \bar{Y}_2}{\frac{1}{q_v} + \frac{n_2}{(\mathbf{t}^2)^{(j+1)}}}, \left(\frac{1}{q_v} + \frac{n_2}{(\mathbf{t}^2)^{(j+1)}}\right)^{-1}\right)$, where $\bar{Y}_2 = \sum_{i=1}^{n_2} \left((Y_2^{(j)})_i - X_i^{(2)} \mathbf{g}^{(j+1)} \right) / n_2$.

(5) Conditional on $z^{(j)}$, $\mathbf{g}^{(j+1)}$, $\mathbf{s}^{(j+1)}$, $\mathbf{t}^{(j+1)}$, $\mathbf{q}^{(j+1)}$, draw for $p^{(j+1)}$ from $\text{Beta}(\mathbf{a}_1 + n_1, \mathbf{a}_2 + n_2)$. With $\mathbf{a}_1 = \mathbf{a}_2 = 1$, little weight will be placed on the prior.

(6) Conditional on $\mathbf{g}^{(j+1)}$, $\mathbf{s}^{(j+1)}$, $\mathbf{t}^{(j+1)}$, $\mathbf{q}^{(j+1)}$, $p^{(j+1)}$, draw for $z^{(j+1)}$ as follows.

Create $p_{it,k} = \frac{p_k^{(j+1)} f_k(\tilde{Y}_{it} | X_{it}, \mathbf{g}^{(j+1)}, \mathbf{s}^{(j+1)}, \mathbf{t}^{(j+1)}, \mathbf{q}^{(j+1)})}{\sum_{k=1,2} p_k^{(j+1)} f_k(\tilde{Y}_{it} | X_{it}, \mathbf{g}^{(j+1)}, \mathbf{s}^{(j+1)}, \mathbf{t}^{(j+1)}, \mathbf{q}^{(j+1)})}$, where $f_1(\cdot)$ and $f_2(\cdot)$ are the

densities for each observation of the first and second components of the mixture as given above in (M), and $p_k^{(j+1)} = p^{(j+1)}$ ($= 1 - p^{(j+1)}$) if $k=1$ (if $k=2$). For each it , draw $u \sim \text{Uniform}$, and set $z_{it}^{(j+1)} = 1$ if $u \leq p_{it,1}$, and 0 otherwise.

From an arbitrary starting value, this is iterated 2,000 times, producing $(z^{(j)}, \mathbf{g}^{(j)}, \mathbf{s}^{(j)}, \mathbf{t}^{(j)}, \mathbf{q}^{(j)}, p^{(j)})$, $j=1,2,\dots,2,000$. The first 500 iterations are discarded, leaving 1,500 draws from posterior distribution of the parameters.

A.4 The Predictive Distribution

Consider the (I+1)st individual, who is exchangeable with (and may, or may not, have been one of) the individuals in the original sample. Denote by XI+1 the exogenous covariates corresponding to XI+1,1,..., XI+1,13, so that it includes pre-treatment covariates and earnings, and the entire set of period indicators for periods 1 to 13 (defined in Section 4.1), but does not include earnings information in post-treatment periods. As well, XI+1 includes a treatment indicator, which we take to be 1 (or 0) to imagine I+1 as part of the treatment group (control group).

Taking both the probit and mixture models together, we wish to compute the joint predictive density of earnings across 13 periods of post-experimental time. Thus, we wish to compute:

$$p(Y_{I+1,1}, Y_{I+1,2}, \dots, Y_{I+1,13} | Data, X_{I+1}) = \int \{ p(Y_{I+1,1} | X_{I+1}, \Theta) p(Y_{I+1,2} | Y_{I+1,1}, X_{I+1}, \Theta) \cdots p(Y_{I+1,13} | Y_{I+1,12}, X_{I+1}, \Theta) \} p(\Theta | Data) d\Theta,$$

where $\Theta = \{\beta, \gamma, \theta, \sigma, \tau, \rho\}$. Conditional on parameters, we use the likelihoods defined in (P) and (M):

$$p(Y_{I+1,t} | Y_{I+1,t-1}, X_{I+1}, \Theta) = \begin{cases} \Pr(Y_{I+1,t} = 0 | Y_{I+1,t-1}^*, X_{I+1}, \Theta) & \text{for } Y_{I+1,t} = 0 \\ \Pr(Y_{I+1,t} > 0 | Y_{I+1,t-1}^*, X_{I+1}, \Theta) \\ \quad \times p(Y_{I+1,t} | Y_{I+1,t-1}^*, X_{I+1}, \Theta) & \text{for } Y_{I+1,t} > 0, \end{cases}$$

where the probability of positive earnings corresponds to the probit model and the density for positive earnings to the mixture model. Finally, we integrate out for the unknown parameters using their posterior distribution.

References

- Albert, J. and S. Chib (1993). "Bayesian Analysis of Binary and Polychotomous Response Data," *Journal of the American Statistical Association*, 88, 669-679.
- Ashenfelter, O., and David Card (1985). "Using the Longitudinal Structure of Earnings to Estimate the Effect of Training Programs," *Review of Economics and Statistics*, 67, 648-660.
- Barberis, Nicholas (1996). "Investing for the Long Run when Returns are Predictable," Harvard University, unpublished.
- Burtless, Gary (1995). "The Case for Randomized Field Trials in Economic and Policy Research," *Journal of Economic Perspectives*, 9, 63-84.
- Caballero, Ricardo (1991). "Earnings Uncertainty and Aggregate Wealth Accumulation," *American Economic Review*, 81, 859-871.
- Chamberlain, Gary, and Guido Imbens (1996). "Hierarchical Bayes Models with Many Instrumental Variables," Harvard Institute of Economic Research, Paper Number 1781.
- Chib, Siddhartha (1992). "Bayes Inference in the Tobit Censored Regression Model," *Journal of Econometrics*, 51, 79-99.
- and Edward Greenberg (1994). "Markov Chain Monte Carlo Simulation Methods in Econometrics," Washington University, manuscript.
- Deaton, Angus, and John Muellbauer (1980). *Economics and Consumer Behavior*. Cambridge: Cambridge University Press.
- Dehejia, Rajeev (1997). "A Decision-Theoretic Approach to Program Evaluation," Chapter 2, Ph.D. Dissertation, Harvard University, unpublished.
- and Sadek Wahba (1996). "Causal Effects in Non-Experimental Studies: Re-Evaluating the Evaluation of Training Programs," Harvard University, unpublished.
- Friend, Benjamin, and Marshal Blume (1975). "The Demand for Risky Assets," *American Economic Review*, 65, 900-22.
- Fisher, R. (1935). *The Design of Experiments*. London: Oliver and Boyd.
- Gelfland, A.E., and A.F.M. Smith (1990), "Sampling-Based Approaches to Calculating Marginal Densities," *Journal of the American Statistical Association*, 85, 398-409.
- Gelman, Andrew, John Carlin, Hal Stern, and Donald Rubin (1996). *Bayesian Data Analysis*. London: Chapman and Hall.

Geman, S., and D. Geman (1984), "Stochastic Relaxation, Gibbs Distributions, and the Bayesian Restoration of Images," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 6, 721-741.

Geweke, John, and Michael Keane (1996). "An Empirical Analysis of the Male Income Dynamics in the PSID: 1968-1989," University of Minnesota, unpublished.

Greenberg, David, and Michael Wiseman (1992). "What Did the OBRA Demonstrations Do," in Manski and Garfinkel (1992).

Hall, Robert (1988). "Intertemporal Substitution," *Journal of Political Economy*, 96, 339-57.

Heckman, James and Joseph Hotz (1989). "Choosing Among Alternative Nonexperimental Methods for Estimating the Impact of Social Programs: The Case of Manpower Training," *Journal of the American Statistical Association*, 84, 862-74.

----- (1990). "Varieties of Selection Bias," *American Economic Review*, 80, 313-318.

----- (1992). "Evaluating Welfare and Training Programs," in Manski and Garfinkel (1992).

----- and Jeffrey Smith (1995). "Assessing the Case for Social Experiments," *Journal of Economic Perspectives*, 9, 85-110.

----- and ----- (1998). "Evaluating the Welfare State," NBER Working Paper No. 6542.

-----, -----, and Nancy Clements (1997). "Making the Most Out of Programme Evaluations and Social Experiments: Accounting for Heterogeneity in Programme Impacts," *Review of Economic Studies*, 64(4), 487-536.

----- and Richard Robb (1985). "Alternative Methods for Evaluating the Impact of Interventions," in James Heckman and Burton Singer (eds.), *Longitudinal Analysis of Labor Market Data*. Econometric Society Monograph No. 10. Cambridge: Cambridge University Press.

----- and ----- (1986). "Alternative Methods for Solving the Problem of Selection Bias in Evaluating the Impact of Treatment on Outcomes," in Howard Rainer (ed.), *Drawing Inferences from Self-Selected Samples*. New York: Springer-Verlag.

Kandel, Shmuel, and Robert Stambaugh (1996). "On the Predictability of Stock Returns: An Asset-Allocation Perspective," *Journal of Finance*, LI, 385-424.

Lalonde, Robert (1986). "Evaluating the Econometric Evaluations of Training Programs," *American Economic Review*, 76, 604-620.

Manski, Charles (1989). "Anatomy of the Selection Problem," *Journal of Human Resources*, 24, 343-360.

----- (1993). "The Selection Problem," in C. Sims (ed.), *Advances in Econometrics*. Cambridge: Cambridge University Press.

----- (1995). "The Mixing Problem in Program Evaluation," *Identification Problems in the Social Sciences*, Chapter 3. Cambridge: Harvard University Press.

----- (1996). "Learning about Treatment Effects from Experiments with Random Assignment of Treatments," *Journal of Human Resources*, 31, 709-733.

----- and Irwin Garfinkel (1992). *Evaluating Welfare and Training Programs*. Cambridge: Harvard University Press.

Neyman, J. (with K. Iwazskiewicz and S. Kolodziejczyk) (1935), "Statistical Problems in Agricultural Experimentation" (with discussion), *Supplement of Journal of the Royal Statistical Society*, 2, 107-180.

Robert, Christian (1996). "Mixtures of Distributions: Inference and Estimation," in W.R. Gilks, S. Richardson, and D.J. Spiegelhalter (eds.), *Markov Chain Monte Carlo in Practice*, London: Chapman and Hall.

Rubin, Donald (1977). "Assignment to a Treatment Group on the Basis of a Covariate," *Journal of Educational Statistics*, 2, 1-26.

----- (1978). "Bayesian Inference for Causal Effects: The Role of Randomization," *The Annals of Statistics*, 6, 34-58.

Riccio, James, Daniel Friedlander, and Stephen Freedman (1994). *GAIN: Benefits, Costs, and Three-Year Impacts of a Welfare-to-Work Program*. New York: Manpower Demonstration Research Corporation.

Rossi, Peter, Robert McCulloch, and Greg Allenby (1995). "Hierarchical Modelling of Consumer Heterogeneity: An Application to Target Marketing," in C. Gatsonis, J. Hodges, R. Kass, and N. Singpurwalla (eds.), *Case Studies in Bayesian Statistics*, Volume II, *Lecture Notes in Statistics*, 105. New York: Springer-Verlag.

Tanner, M., and W. Wong (1987). "The Calculation of Posterior Distributions by Data Augmentation," *Journal of the American Statistical Association*, 82, 528-550.

Data Legend

Variable	Description
CHILD4	Number of children less than age 4
CHILD45	Number of children between ages 4 and 5
CHILD611	Number of children between ages 6 and 11
CHILD18	Number of children between ages 12 and 18
CHILD19	Number of children aged 19 and greater
CAREAD	Score on reading test
CAMATH	Score on mathematics test
GRADE	Educational attainment (grade 0 to 20)
HRWAGE	Most recently recorded hourly wage
FAM.TYPE	Indicator for households with single head
AGE	Age
EXPER.	Indicator for experimental unit
CONTROL	Indicator for control unit
SEXF	Indicator for female participants
REFUGEE	Indicator of refugee status
CUR.AFDC	Indicator for receiving AFDC in pre-experimental time
PREVTR	Indicator for previous training or job search activities
ETH. WHITE	Ethnicity Indicator, White
ETH. HISP.	Ethnicity Indicator, Hispanic
ETH.BLACK	Ethnicity Indicator, Black
ETH.NATIVE	Ethnicity Indicator, Native Indian
ETH.IND.CH.	Ethnicity Indicator, Indo-Chinese
ETH.OTH.AS	Ethnicity Indicator, Other Asian
ETH.PACF.	Ethnicity Indicator, Pacific Islander
ETH.FILIP.	Ethnicity Indicator, Filipino
ETHOTHR	Ethnicity Indicator, Other
AVG.UNEMP	Average county unemployment rate, at registration
PEARN _x	Earnings in quarter <i>x</i> of pre-experimental time
PEARN _{xZ}	Indicator of zero earnings, pre-experimental quarter <i>x</i>
EARN _x	Earnings in quarter <i>x</i> of post-experimental time
EARN _{xZ}	Indicator of zero earnings, post-experimental quarter <i>x</i>
PAFDC _x	AFDC receipts, pre-experimental quarter <i>x</i>
AFDC _x	AFDC receipts, post-experimental quarter <i>x</i>
PFDSTMP _x	Food Stamps receipts, pre-experimental quarter <i>x</i>
FDSTMP _x	Food Stamps receipts, post-experimental quarter <i>x</i>
EMPPQ _x	Employment status, pre-experimental quarter <i>x</i>
EMPQ _x	Employment status, post-experimental quarter <i>x</i>

Table 1: The Sample

	Alameda	Butte	Los Angeles	Riverside	San Diego	Tulare
GAIN:						
Treated Group	685	1717	3730	5808	8711	2693
Control Group	682	458	2124	1706	1810	1146
Total	1367	2175	5854	7514	10521	3839
AFDC:						
Total	30305	5663	231356	24000	50958	14673

Notes: The GAIN sample sizes are from the public use file of the GAIN data. The AFDC total represents the number of AFDC cases (both single-parent and two-parent households) in the six evaluation counties in December 1990 (see Riccio, *et al.* (1994), Table 1.1).

Table 2: Data Description, Alameda County

Variable	Mean	Standard deviation
CHILD4	0.19	0.49
CHILD45	0.23	0.46
CHILD611	1.16	4.68
CHILD18	0.88	2.29
CHILD19	0.25	0.60
CAREAD	206.27	98.00
CAMATH	192.44	94.96
GRADE	10.79	3.02
HRWAGE	3.74	2.73
FAM.TYPE	0.62	
AGE	35.39	8.85
EXPER.	0.50	
SEXF	0.86	
REFUGEE	0.09	
CUR.AFDC.	0.99	
PREVTR	0.24	
ETH.WHITE	0.18	
ETH.HISP.	0.08	
PEARN10	165.02	740.14
PEARN9	153.17	675.96
PEARN8	154.53	747.70
PEARN7	187.67	1036.91
PEARN6	156.83	615.03
PEARN5	170.37	771.74
PEARN4	185.30	726.89
PEARN3	151.60	685.37
PEARN2	153.64	642.86
PEARN1	167.17	714.04
PEARNZ10	0.87	
PEARNZ9	0.88	
PEARNZ8	0.87	
PEARNZ7	0.87	
PEARNZ6	0.87	
PEARNZ5	0.87	
PEARNZ4	0.87	
PEARNZ3	0.87	
PEARNZ2	0.87	
PEARNZ1	0.87	

Table 3: Treatment Effect on Probability of Unemployment

Post-experimental period	Treatment effect	Standard error
1	0.024	0.017
2	-0.011	0.019
3	-0.019	0.020
4	-0.034	0.021
5	-0.044	0.021
6	-0.018	0.021
7	-0.021	0.022
8	-0.048	0.022
9	-0.062	0.023
10	-0.061	0.023
11	-0.044	0.023
12	-0.027	0.022
13	-0.063	0.022

Note: A probit is used; covariates include variables for the number of children (CHILD4-CHILD19), reading and writing test scores, grade, age, sex, ethnicity, and earnings histories (PEARN10-PEARN1). The treatment effect is computed as the discrete difference between the probability of unemployment with the treatment indicator set to 0 and 1, where the value of other covariates is set to their sample mean. The delta method is used to compute standard errors.

Table 4: Regression Coefficients of Treatment Indicator for Post-experimental Earnings

Post-experimental period	OLS treatment effect, with covariates	Standard error
1	-47.6	22.8
2	-9.2	39.2
3	35.1	45.9
4	67.6	56.8
5	111.3	54.1
6	85.0	61.1
7	84.8	66.5
8	95.3	68.6
9	203.1	76.0
10	232.1	79.5
11	194.5	86.4
12	150.7	88.8
13	206.6	90.8

Table 5a: Characteristics of an Individual (“Ms. Thirteen Fifty-Three”)

CHILD4	CHILD45	CHILD611	CHILD18	CHILD19	CAREAD	CAMATH	GRADE	HRWGE	FAM.TYPE
0	0	0	1	0	241	225	15	8	30
AGE	SEXF	REFUGEE	ETH.WHT	ETH.HSP	AVG.UNEM	PEARN10	PEARN9	PEARN8	PEARN7
42	1	0	1	0	4.1	0	0	0	3667
PEARN6	PEARN5	PEARN4	PEARN3	PEARN2	PEARN1	PAFDC7	PAFDC6	PAFDC5	PAFDC4
425	0	0	0	0	0	1605	1605	1551	1443
PAFDC3	PAFDC2	PAFDC1							
1518	1518	1518							

Table 5b: Characteristics of an Individual (“Ms. One”)

CHILD4	CHILD45	CHILD611	CHILD18	CHILD19	CAREAD	CAMATH	GRADE	HRWGE	FAM.TYPE
0	0	1	2	0	221	191	9	2.5	30
AGE	SEXF	REFUGEE	ETH.WHT	ETH.HSP	AVG.UNEM	PEARN10	PEARN9	PEARN8	PEARN7
32	1	0	0	1	3.07	0	0	0	0
PEARN6	PEARN5	PEARN4	PEARN3	PEARN2	PEARN1	PAFDC7	PAFDC6	PAFDC5	PAFDC4
0	0	0	0	0	0	2259	2364	2364	2364
PAFDC3	PAFDC2	PAFDC1							
2364	2472	2472							

Table 5c: Characteristics of an Individual (“Ms. Ten”)

CHILD4	CHILD45	CHILD611	CHILD18	CHILD19	CAREAD	CAMATH	GRADE	HRWGE	FAM.TYPE
0	0	2	0	0	223	240	10	5.85	30
AGE	SEXF	REFUGEE	ETH.WHT	ETH.HSP	AVG.UNEM	PEARN10	PEARN9	PEARN8	PEARN7
27	1	0	1	0	3.7	0	0	0	0
PEARN6	PEARN5	PEARN4	PEARN3	PEARN2	PEARN1	PAFDC7	PAFDC6	PAFDC5	PAFDC4
0	0	0	0	0	0	1899	1899	1926	1989
PAFDC3	PAFDC2	PAFDC1							
1989	1989	2082							

Table 6a: Mean and Variance of Predicted Earnings, Ms. Thirteen Fifty-Three

Post-treatment earnings period	Treatment			Control		
	Probability of positive earnings	Mean post-treatment earnings	Standard deviation	Probability of positive earnings	Mean post-treatment earnings	Standard deviation
1	0.12	21	96	0.05	38	294
2	0.26	35	135	0.10	37	201
3	0.36	215	418	0.12	246	900
4	0.47	360	587	0.15	342	1358
5	0.53	443	659	0.15	367	1291
6	0.55	536	985	0.18	430	1258
7	0.59	590	797	0.22	596	1618
8	0.62	650	820	0.21	640	1924
9	0.66	301	1190	0.22	202	697
10	0.67	276	567	0.23	277	1291
11	0.68	890	1025	0.23	893	2490
12	0.67	882	1049	0.22	765	2097
13	0.69	368	900	0.18	285	1432

Table 6b: Predicted Earnings, With and Without Uncertainty, Ms. Ten

Post-treatment earnings period	Ignoring parameter uncertainty				Accounting for parameter uncertainty			
	Treated		Control		Treated		Control	
	Predicted earnings	Standard deviation	Predicted earnings	Standard deviation	Predicted earnings	Standard deviation	Predicted earnings	Standard deviation
1	29	217	51	245	23	220	40	233
2	76	356	131	581	77	340	119	502
3	208	851	207	701	179	718	227	772
4	316	1014	292	994	302	1127	285	955
5	349	1124	309	986	260	1316	329	1085
6	429	1328	322	1020	376	1277	384	1141
7	443	1427	399	1534	491	1657	417	1330
8	162	1498	200	1116	581	1715	428	1222
9	638	1710	383	1272	640	1857	463	1385
10	674	1813	533	1537	639	2058	478	1448
11	844	1880	514	1501	679	1882	575	1709
12	641	1870	611	1755	604	1912	589	1716
13	898	2570	628	1860	738	2062	484	1836
Log utility*	48.85		48.53		48.72		48.75	
CRRA (3)*	4.86607		4.86606		4.86606		4.86607	

Note: * For expected utility comparisons, \$100 is added to the distribution of treated and control earnings.

Table 7: Groups Benefiting the Most and Least from GAIN, Comparing Mean Covariates

Variable	Those who prefer GAIN ^a	Those who prefer AFDC ^b
Number	571	789
Avg. Earnings GAIN	854	252
Avg. Earnings AFDC	531	273
CHILD4	0.16	0.21
CHILD45	0.23	0.24
CHILD611	0.94	0.85
CHILD18	0.56	1.02
CHILD19	0.09	0.36
CAREAD	211.99	201.95
CAMATH	197.33	188.69
GRADE	11.85	10.03
HRWAGE	5.05	3.56
FAM.TYPE	30.47	30.74
AGE	32.43	37.50
EXPER.	0.47	0.53
CONTROL	0.53	0.47
SEXF	0.90	0.82
REFUGEE	0.06	0.11
CUR.AFDC	1.00	0.98
PREVTR	0.46	0.08
ETH. WHITE	0.19	0.17
ETH. HISP.	0.05	0.10
ETH.BLACK	0.69	0.60
ETH.NATIVE	0.00	0.01
ETH.IND.CH.	0.05	0.09
ETH.OTH.AS	0.01	0.02
ETH.PACF.	0	0.00
ETH.FILIP.	0.01	0.01
ETHOTHR	0.02	0.04
AVG.UNEMP	3.95	4.05
PEARN10	323.02	52.14
PEARN9	281.42	61.71
PEARN8	296.80	52.93
PEARN7	399.16	36.29
PEARN6	274.97	72.71
PEARN5	362.57	32.78
PEARN4	407.32	26.27
PEARN3	333.22	20.75
PEARN2	299.38	47.28
PEARN1	268.82	95.10

Notes: (a) Individuals whose expected utility is higher from total earnings under GAIN than AFDC, where preferences are CRRA, with relative risk aversion equal to 3.

(b) The complement of (a).

Table 8: Comparing GAIN and AFDC, Average of Outcomes Per Person^a

Policy	Labor earnings per quarter	Total income per quarter	Probability of employment	Total fiscal expenditures	Earnings net-of-costs per quarter ^b
GAIN	463	2,425	0.2042	1,883	202
AFDC	372	2,359	0.1843	1,636	387
Standard error	56	67	0.018	85	-48

Notes: (a) Means are computed from the empirical distribution.

(b) Costs are normalized to zero for AFDC, and are an additional \$3638 for 13 quarters of GAIN.

Table 9a: Social Welfare Comparisons for Five Different Policies, Alameda

Policy	SWF(1)	SWF(2)	SWF(3)	SFW(4)	SWF(5)
GAIN	530 [458,622]	2640 [2565,2750]	0.2115	5755 [5746,5772]	-165 [-182,-134]
AFDC	405 [350,458]	2506 [2454,2555]	0.1788	2101 [2098,2105]	0 [0,0]
Mandated [Treated=707]	624 [548,716]	2728 [2650,2830]	0.2369	4001 [3992,4018]	70 [56,108]
Choice (1) [Treated=396]	654 [579,756]	2756 [2678,2855]	0.2218	3162 [3153,3178]	167 [147,216]
Choice (2) [Treated=463]	651 [577,752]	2754 [2677,2851]	0.2268	3341 [3332,3357]	150 [133,198]
Choice (3) [Treated=571]	639 [567,746]	2744 [2668,2847]	0.2320	3635 [3626,3652]	113 [103,168]

Notes: Each set of values of the parameters from the posterior distribution defines a state of the world. For each state of the world the social welfare functions are computed. Thus, there is a distribution of these SWFs over the various states of the world. These are summarized by the mean and the 2.5 and 97.5 percentiles of the distributions.

SWF(1): Average earnings per individual.

SWF(2): Total expected income (earnings + AFDC receipts + food stamp receipts).

SWF(3): Expected proportion of individuals employed (i.e., greater than zero income), averaged over 13 post-experimental quarters.

SWF(4): Total expected expenditure, which includes AFDC payments, food stamp payments, and, for GAIN, includes administrative costs (in excess of AFDC administration) of \$3,638 per individual for 13 quarters.

SWF(5): Increase in average earnings per person per quarter net of costs, relative to AFDC.

GAIN: Mandatory participation in GAIN.

AFDC: Mandatory participation in AFDC.

Mandated: Individuals are required to join the program that maximizes their chances of being employed.

Choice: Individuals choose between AFDC and GAIN based on expected utilities (or certainty equivalents):

(1): risk-neutral agents, discount factor = 0.95;

(2): risk-averse agents (CRRA, $u(x) = (x)^{(1-q)} - 1 / (1-q)$), $q \approx 1$, discount factor = 0.95;

(3): " , $q=3$, discount factor=0.95.

Table 9b: Considering *Ex Post* Inequality, Quantiles of the Earning Distribution

Policy	0.05	0.25	0.5	0.75	0.95	(0.90-0.10)
GAIN	0 [0,0]	0 [0,0]	39 [5,74]	491 [392,606]	1961 [1677,2251]	1252 [1097,1478]
AFDC	0 [0,0]	0 [0,0]	55 [0,101]	601 [492,704]	1784 [1573,1993]	1312 [1153,1469]
Mandated	0 [0,0]	0 [0,0]	122 [79,185]	702 [587,820]	2156 [1979,2335]	1508 [1378,1694]
Choice (1)	0 [0,0]	0 [0,0]	116 [66,171]	772 [668,892]	2271 [2071,2546]	1616 [1504,1840]
Choice (2)	0 [0,0]	0 [0,0]	121 [80,182]	768 [666,880]	2257 [2050,2522]	1607 [1481,1804]
Choice (3)	0 [0,0]	0 [0,0]	122 [79,187]	738 [644,860]	2206 [1993,2500]	1559 [1426,1757]

Note: Each cell presents the median of the posterior distribution of the percentile, and in parentheses the 5th and 95th posterior percentiles.

Table 9c: Expected Utility Comparisons, Alameda

Policy	SWF(6)*	SWF(7)*	SWF(8)*	SWF(9)*
Risk Neutral:				
GAIN	1.0483	0.5492	0.9998	-1.0357
AFDC	1	1	1	-1
Mandated	1.0817	0.8205	1.0036	-0.9760
Choice (1)	1.0927	1.0114	1.0049	-0.9614
Risk Averse, q=1				
GAIN	0.9892	0.2707	0.9949	-1.1240
AFDC	0.9487	0.1972	0.9936	-1.1593
Mandated	1.0098	0.1972	0.9972	-1.1033
Choice (2)	1.0144	0.2707	0.9977	-1.0896
Risk Averse, q=3				
GAIN	0.9295	0.1994	1.0065	-1.2127
AFDC	0.9057	0.1153	1.0048	-1.3159
Mandated	0.9397	0.1153	1.0075	-1.2384
Choice (3)	0.9411	0.1994	1.0078	-1.1992

Notes: * Expected utilities are normalized.

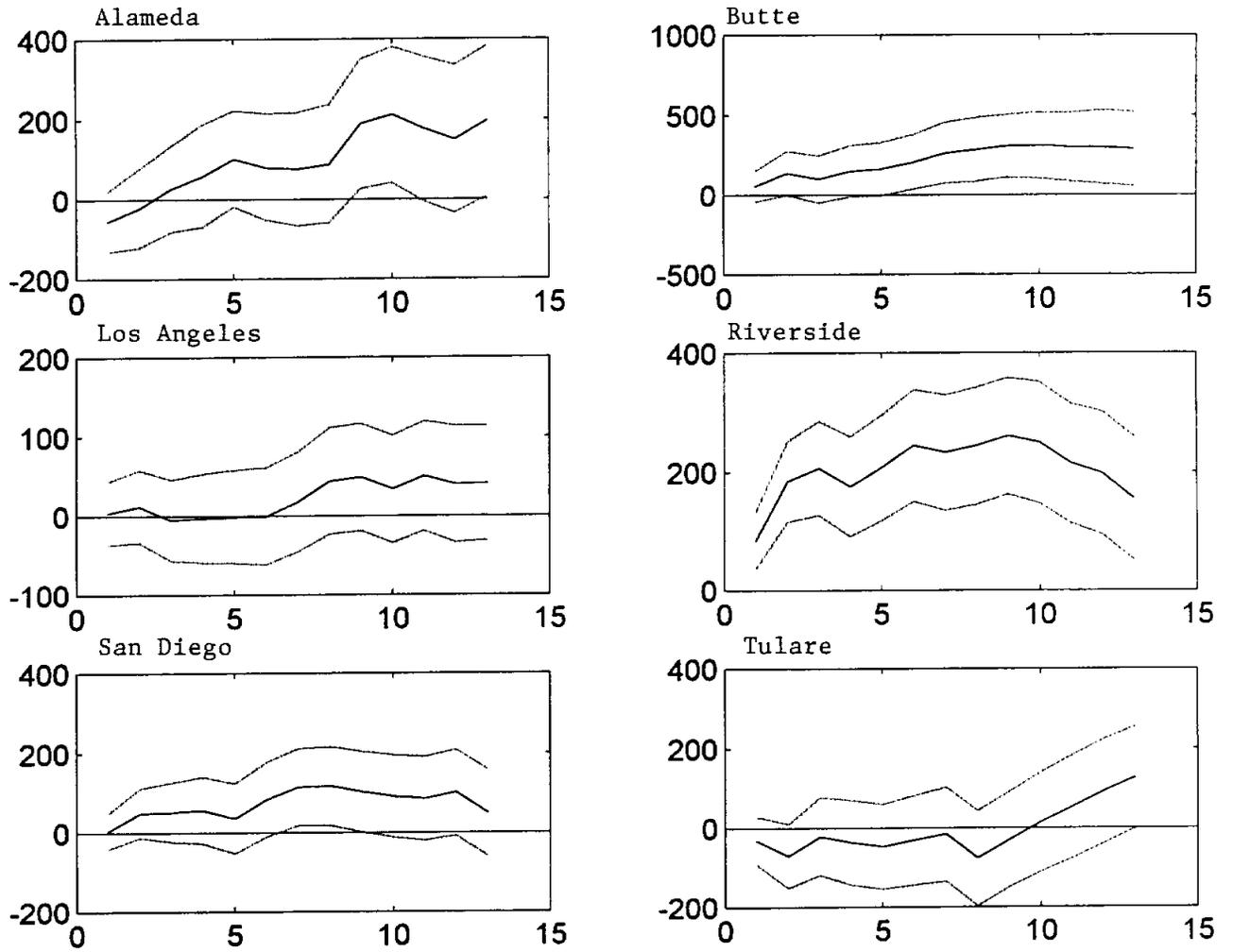
SWF(6): Utilitarian SWF, applied to certainty equivalent of income distribution.

SWF(7): Rawlsian SWF, applied to certainty equivalent of income distribution.

SWF(8): log SWF, applied to certainty equivalent of income distribution.

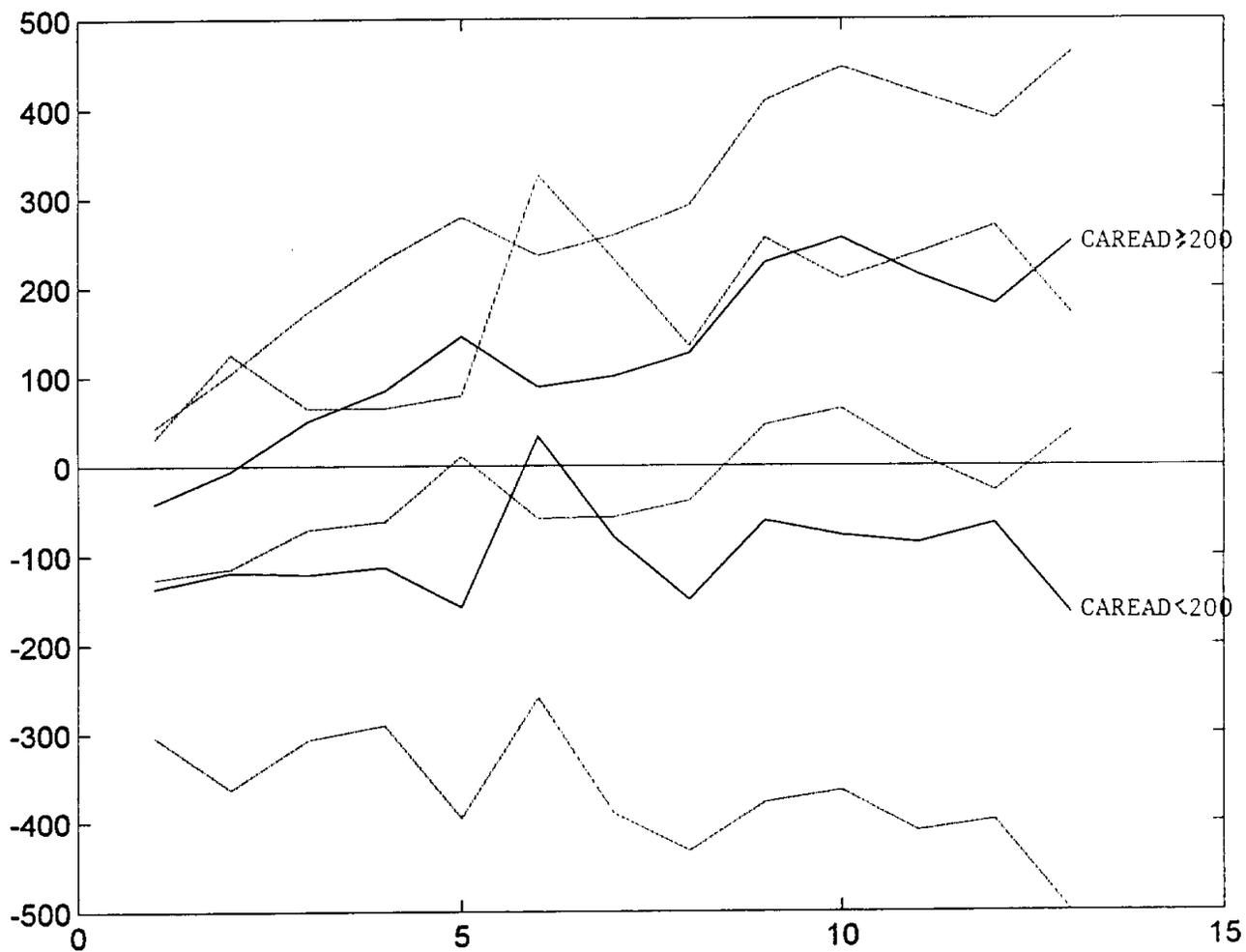
SWF(9): $SWF = \left(\frac{1}{1-\epsilon}\right) \left(\sum_i (u_i)^{1-\epsilon}\right)$, $\epsilon=3$, applied to certainty equivalent of income distribution.

Figure 1: Treatment Effect Six Counties



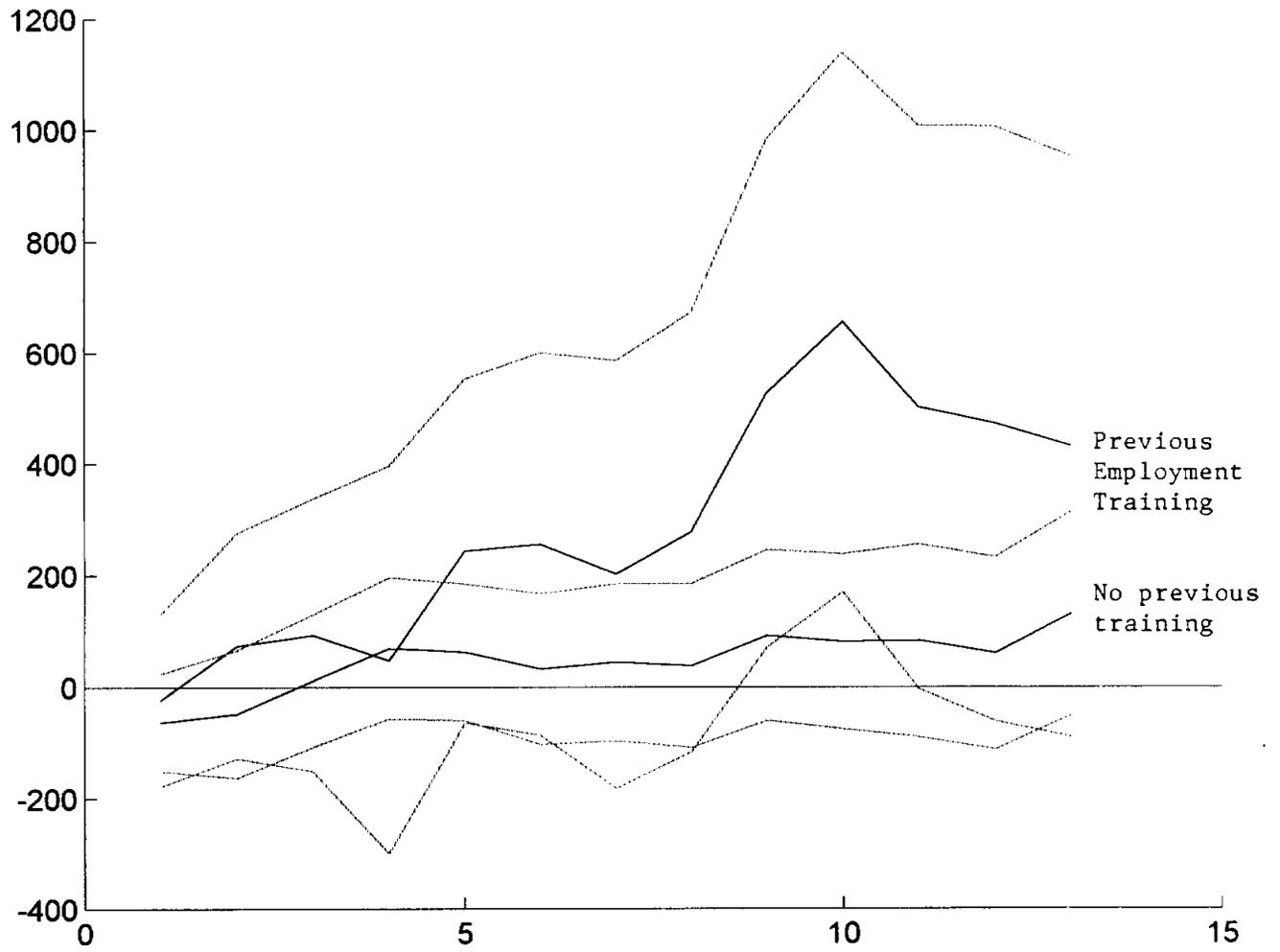
Dashed Lines: +/- two standard errors

Figure 2: Treatment Effects, Alameda



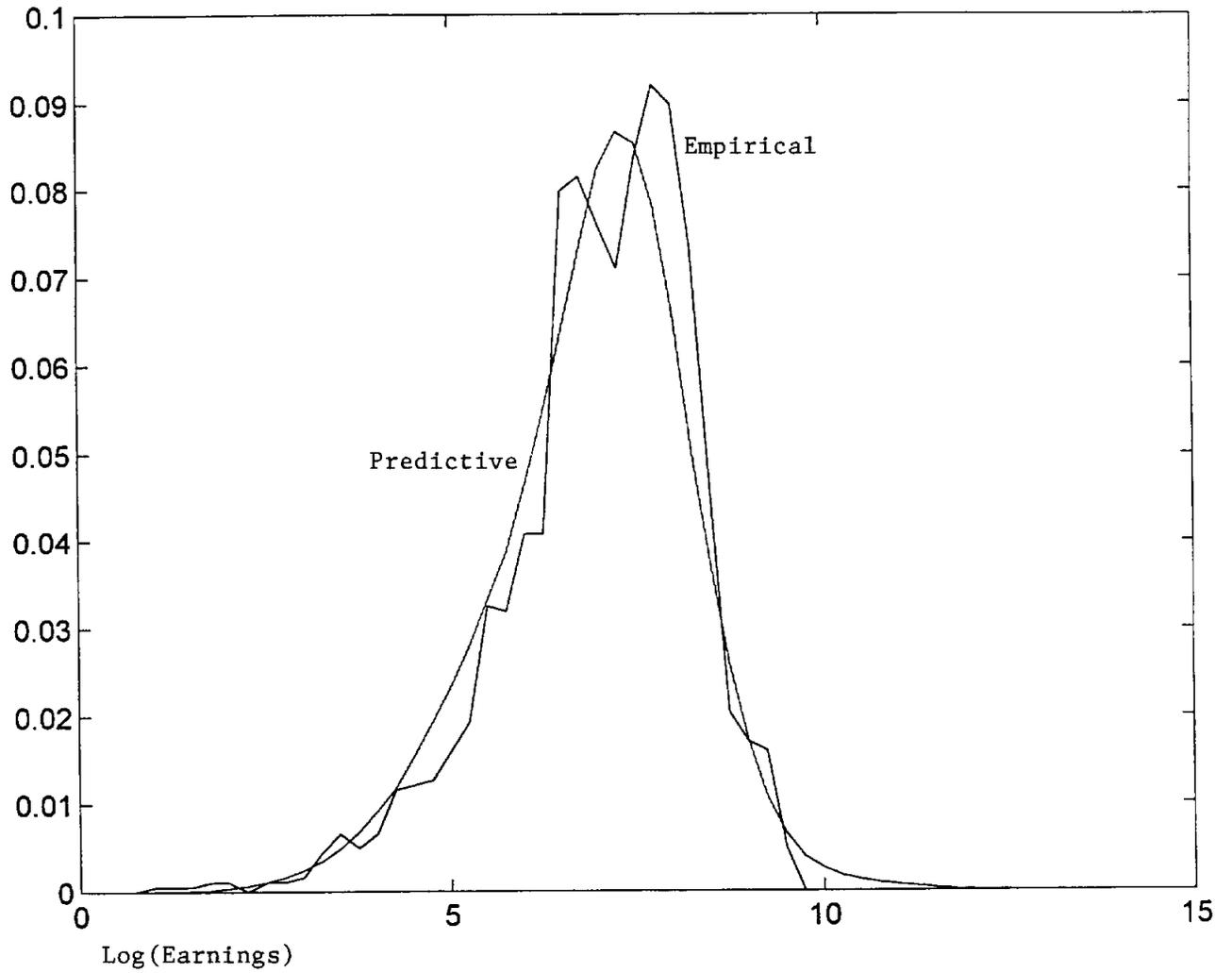
Dashed Lines: +/- two standard errors

Figure 3: Treatment Effects, Alameda



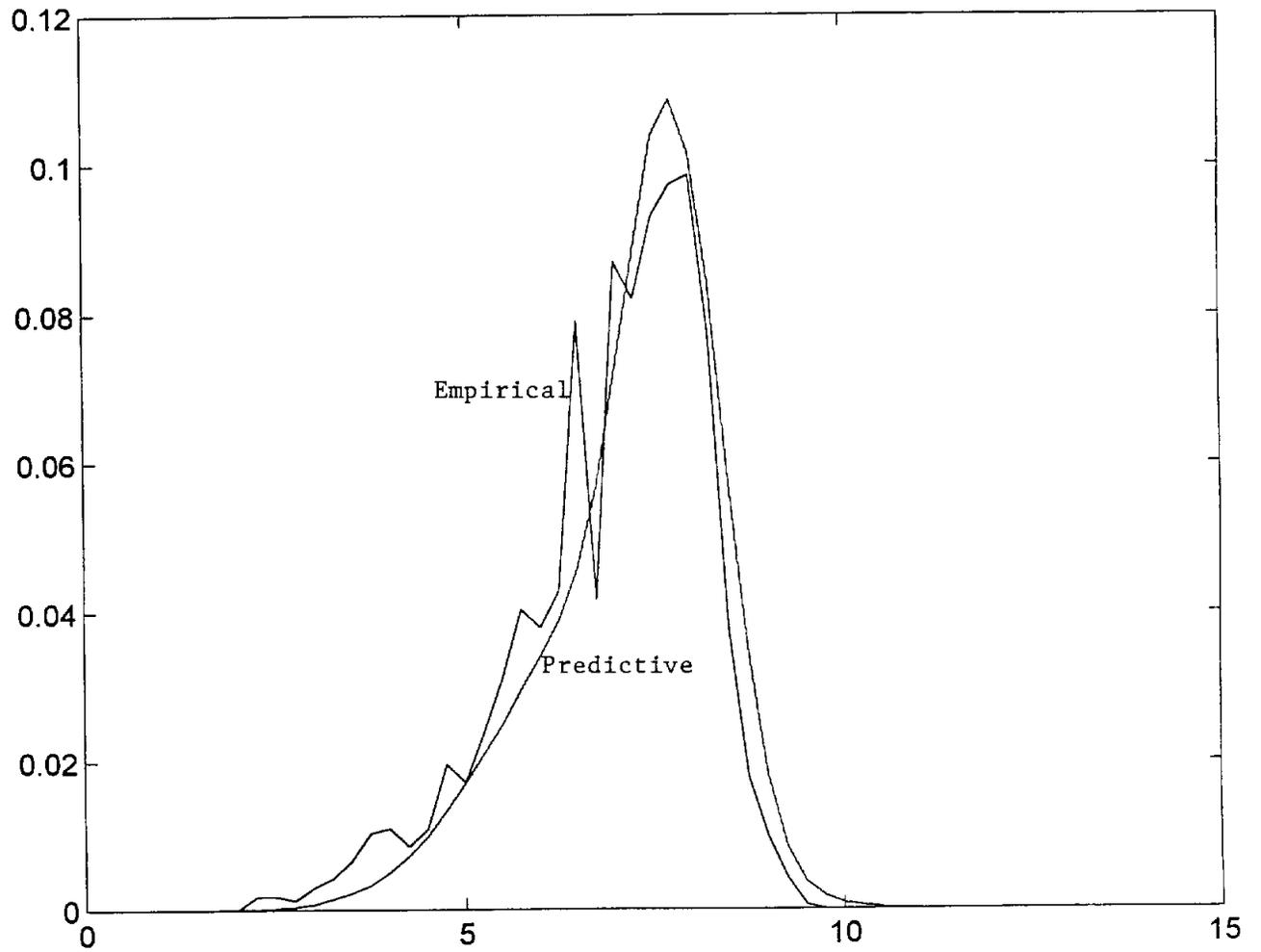
Dashed Lines: +/- two standard errors

Figure 4: Average Post-Treatment Log Earnings
Per Person Per Quarter for Positive Earnings
Empirical and Predictive Densities, Treatment Group



Fraction of Zero Earnings: Empirical 0.80, Predictive 0.79

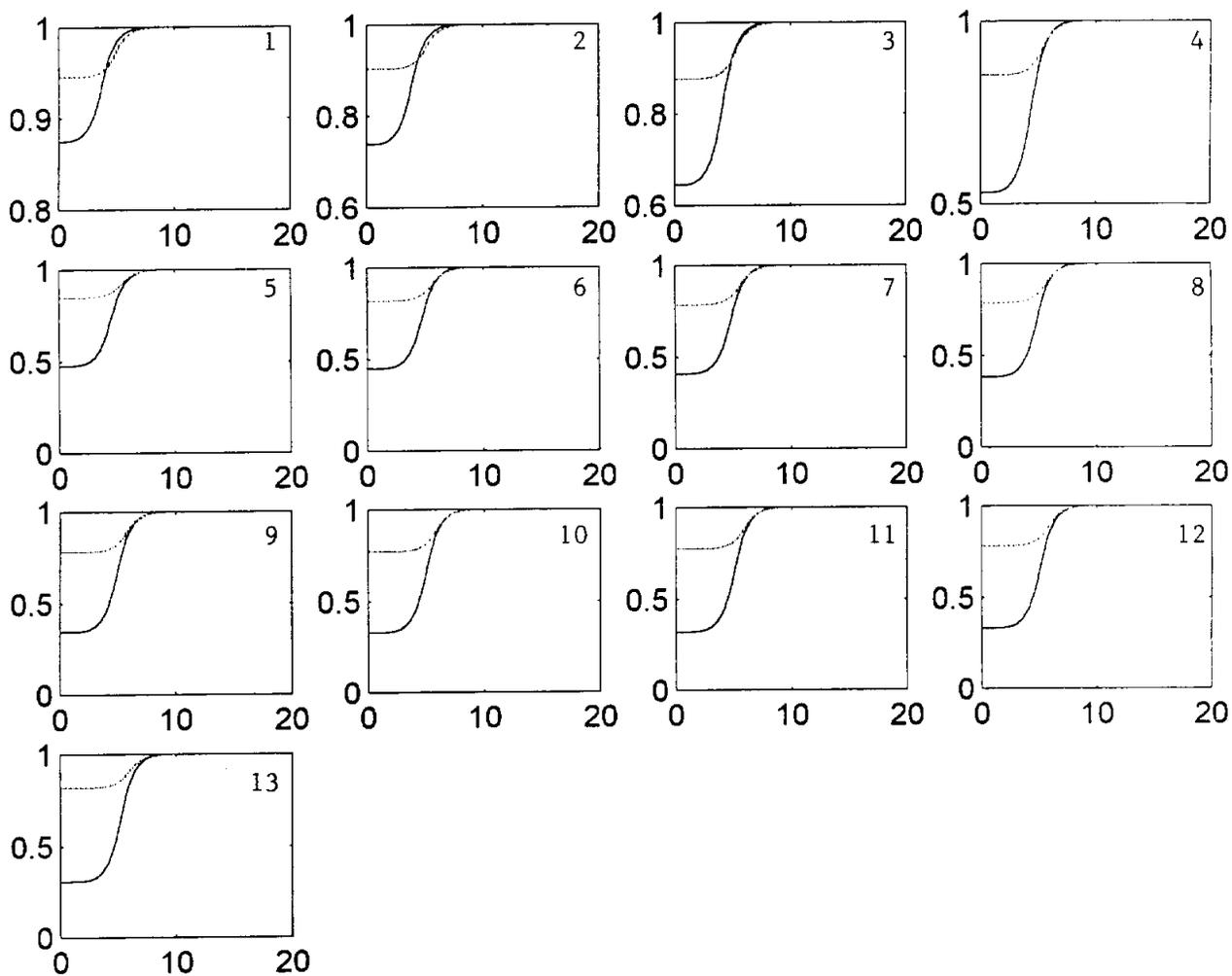
Figure 5: Average Post-Treatment Log Earnings
Per Person Per Quarter for Positive Earnings
Empirical and Predictive Densities, Control Group



Log(Earnings)

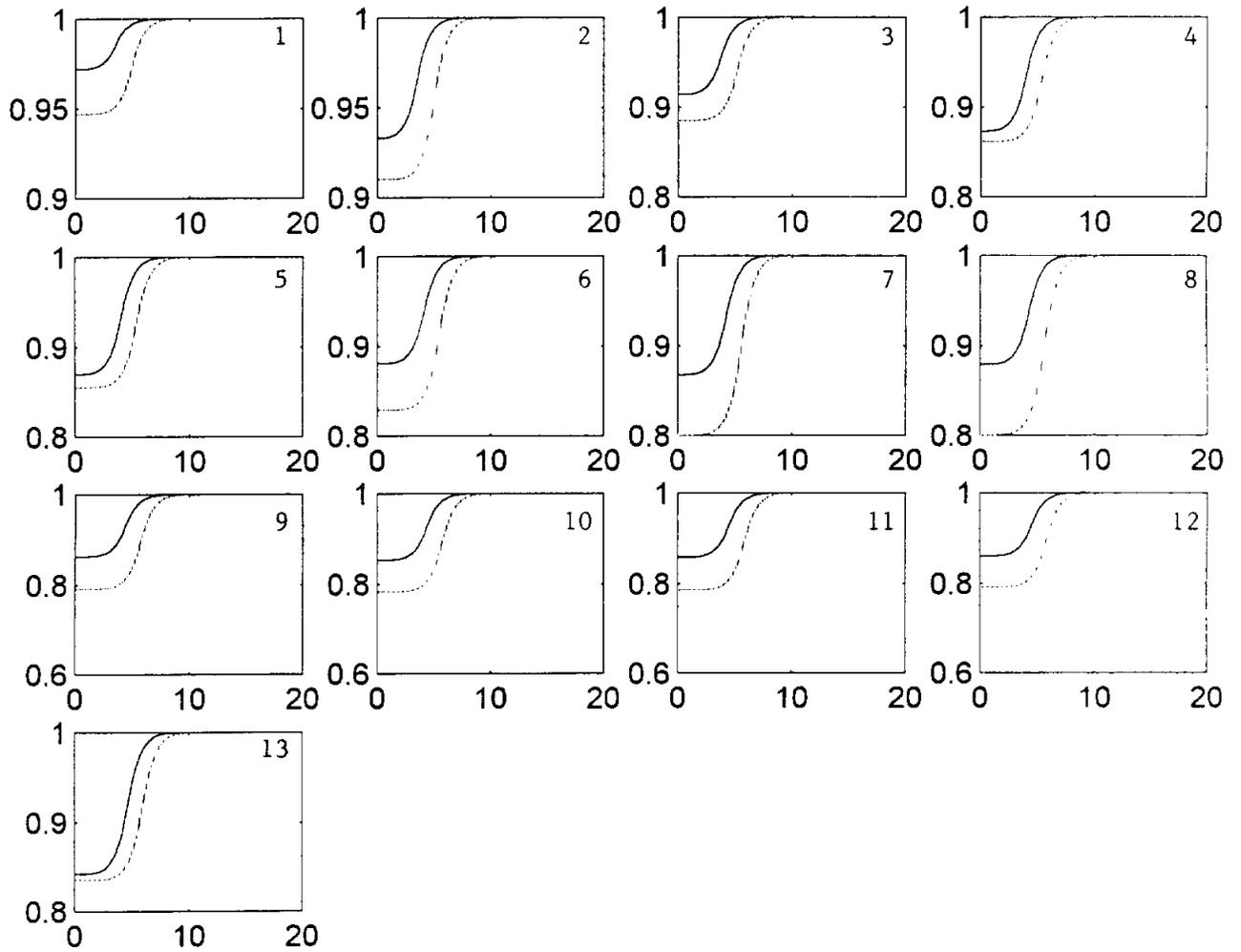
Fraction of Zero Earnings: Empirical 0.82, Predictive 0.82

Figure 6: Cumulative Distribution Functions, Post-Treatment Earnings (Unit 1353)



Log(Earnings+1), Solid=Treated, Dashed=Control

Figure 7: Cumulative Distribution Functions, Post-Treatment Earnings (Unit 1)



Log(Earnings+1), Solid=Treated, Dashed=Control

Figure 8: Cumulative Distribution Functions, SWF(1)

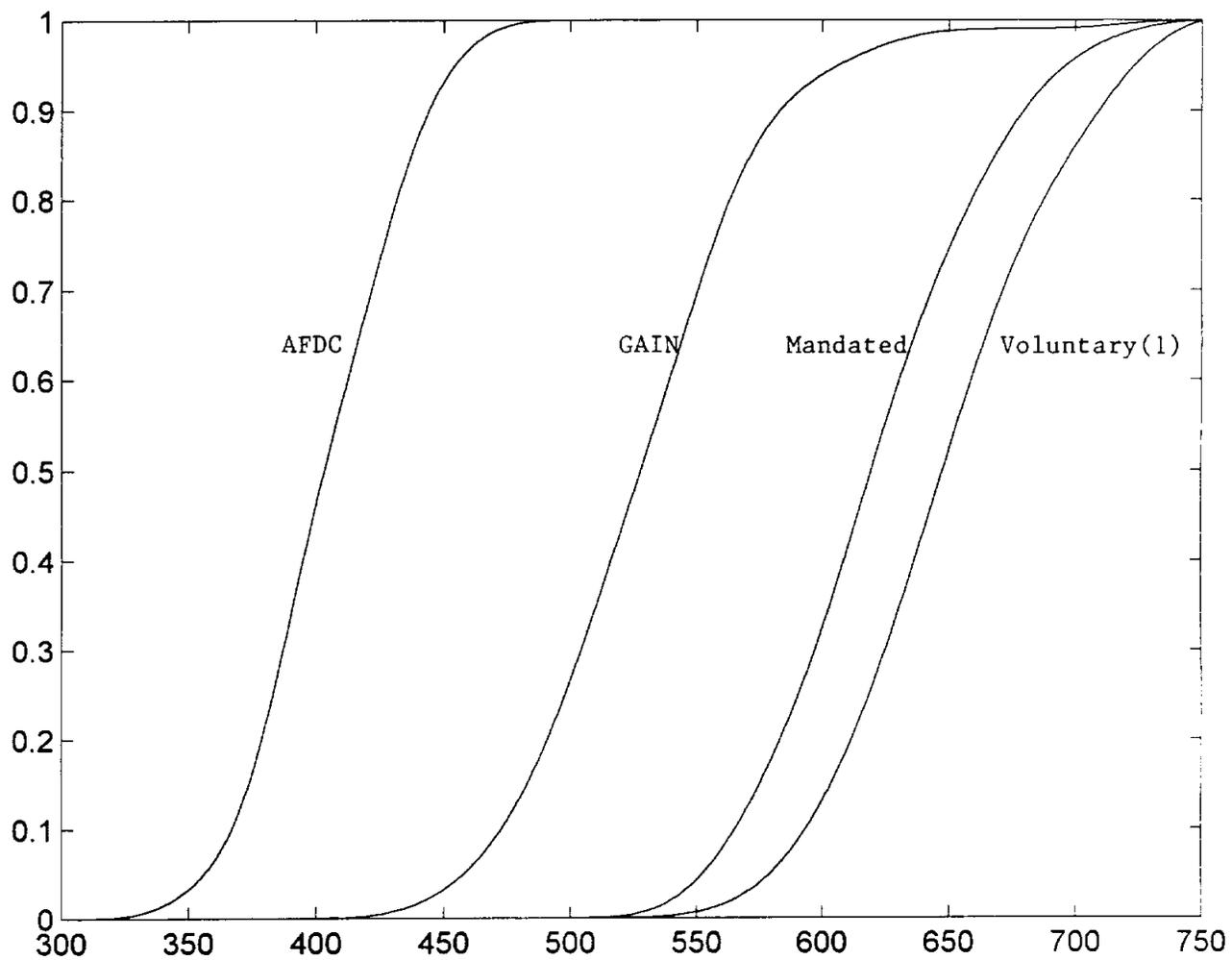


Figure 9: Cumulative Distribution Functions, SWF(5)

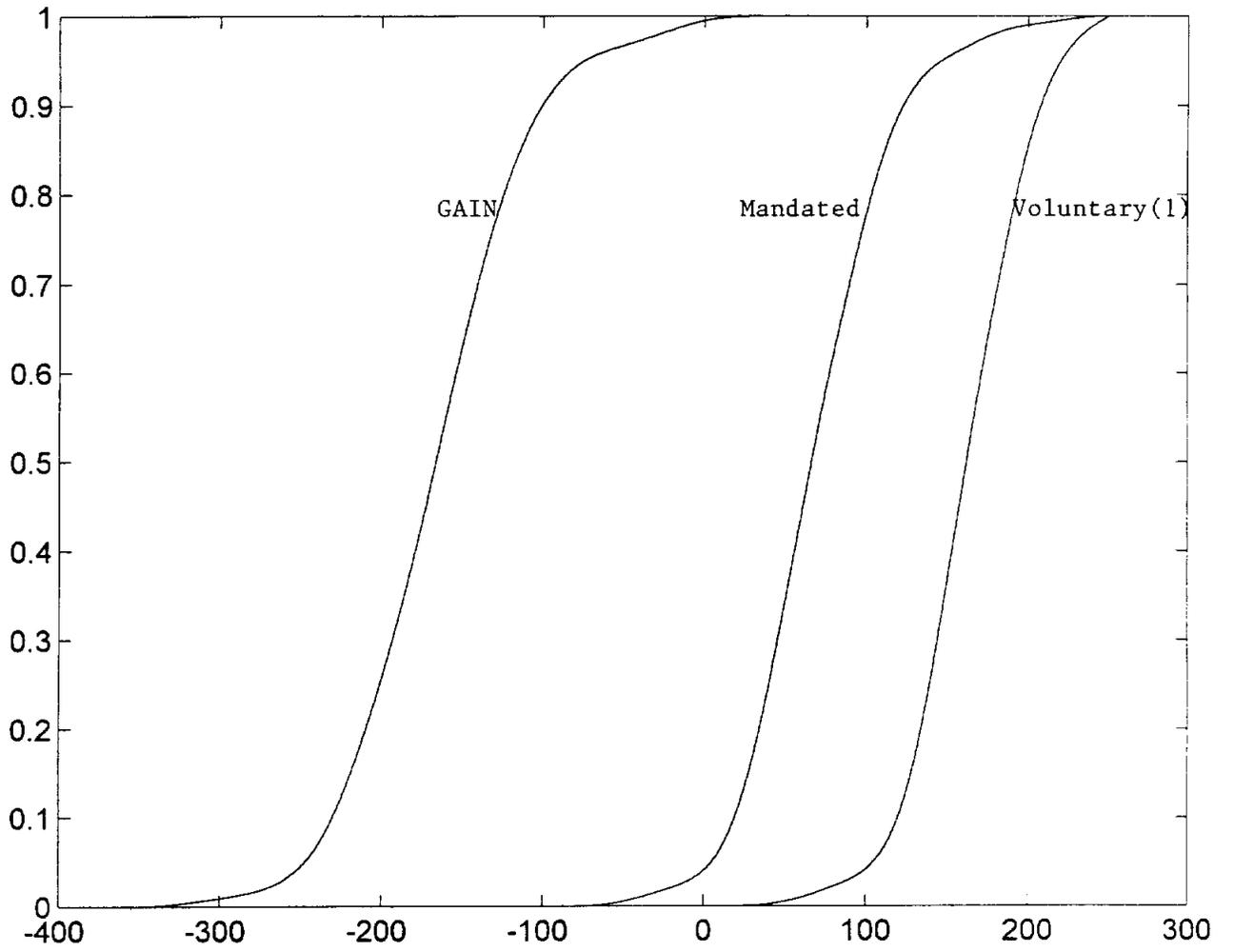


Figure 10: Percentiles of the Earnings Distribution
Average Earnings Per Quarter
(Medians of the Posterior Distributions of the Percentiles)

