What Mean Impacts Miss: Distributional Effects of Welfare Reform Experiments

Preliminary and incomplete, please do not quote without permission.

Marianne P. Bitler
Fellow in Population Studies and the Study of Aging, RAND

Jonah B. Gelbach
University of Maryland (on leave)
RWJ Scholars / University of California at Berkeley

Hilary W. Hoynes
University of California, Davis and NBER*

First version: January 2003
This version: March 26, 2003

Abstract

Standard labor supply theory predicts systematic heterogeneity in the impact of recent welfare reforms on earnings, transfer payments, and total income. Simulation evidence suggests this heterogeneity could be important. In this paper we use experimental data to estimate quantile treatment effects (QTEs) of Connecticut's Jobs First reform. We find considerable evidence of systematic heterogeneity, which is generally consistent with theoretical predictions. Moreover, we find that mean impacts computed for subgroups would not have uncovered a number of important findings. For example, we find that earnings *fall* at the very top of the distribution, as theory predicts should happen when more generous disregards encourage women to stay on or re-enter welfare. We also find that Jobs First has essentially no impact on earnings at the bottom of the earnings distribution. Income at the bottom deciles is constant before time limits hit (contrary to large positive mean impacts for a disadvantaged subgroup) and falls after time limits begin to take effect (again contrary to a zero effect using means). These findings suggest cause for concern about the ability of programs like Jobs First to end welfare dependence of significant numbers of women.

---

*Correspondence to Hoynes at UC Davis, Department of Economics, 1152 Social Sciences and Humanities Building, One Shields Avenue, Davis, CA 95616-8578, phone (530) 752-3226, fax (530) 752-9382, or hoynes@ssds.ucdavis.edu; Gelbach at gelbach@glue.umd.edu; or Bitler at bitler@rand.org. Bitler gratefully acknowledges the financial support of the National Institute of Child Health and Human Development and the National Institute on Aging. This work has not been formally reviewed or edited. The views and conclusions are those of the authors and do not necessarily represent those of the RAND corporation. We are very grateful to MDRC for providing the public access to the experimental data used here. We would also like to thank David Card, Mary Daly, Guido Imbens, and Jeff Smith for helpful conversations.

1 Introduction

More than five years has now passed since the elimination of Aid to Families With Dependent Children (AFDC), the principal U.S. cash assistance program for six decades. In 1996, enactment of the Personal Responsibility and Work Opportunity Reconciliation Act (PRWORA) required all 50 states to replace AFDC with a Temporary Assistance for Needy Families (TANF) program. State TANF programs differ from AFDC in many fundamental ways. Key examples include lifetime limits on program participation, stringent work requirements, financial sanctions for failure to comply with these requirements, and enhanced work incentives through expanded earnings disregards. A critical element in evaluating the performance of this dramatic policy change is measuring the impact of TANF on earnings and family income. As we discuss below, mean impacts of these policies on earnings and income are theoretically indeterminate.

Existing research on the effects of welfare reform on earnings and income shares an important feature—estimation of mean impacts.[1] Several studies use nonexperimental data from the Current Population Survey (CPS) to examine the impact of PRWORA, and the state waivers that preceded it, on income. Evidence from these studies is mixed. For example, Moffitt (1999) finds no impact of waivers on family income, while Grogger (Forthcoming) finds that welfare reform increased mean income for female heads of household. Further, experimental studies examining state reforms implemented before 1996 via TANF-like waivers show that generous increases in the earnings disregards are important for generating mean income gains. However, these gains disappear after time limits (Bloom & Michalopoulos (2001), Grogger et al. (2002)).

In this paper, we shift attention away from means. Instead we allow for the possibility of detecting heterogeneous impacts of welfare reform by estimating quantile treatment effects (QTE). We use public-use data files from three randomized experiments, which were conducted during the mid- to late-1990s to evaluate state waivers from AFDC rules. We focus attention on Connecticut's Jobs First waiver because this program involved both a radical increase in the earnings disregard and a very short (21-month) time limit. In future versions, we will also present results from Florida's Family Transition Program (FTP) and the Minnesota Family Investment Program (MFIP), each of

[1] The welfare reform literature that has developed in the last several years is enormous. We confine our discussion of this literature to a few papers having particular relevance to our study of income and heterogeneity. For comprehensive summaries of this research, see the excellent reviews by Blank (2002), Moffitt (2002), and Grogger, Karoly & Klerman (2002).

which incorporated a smaller, though still significant, increase in the earnings disregard. FTP also had a fairly short time limit, while MFIP had a strict work requirement. These programs embody the key features of TANF, and the Jobs First and FTP demonstrations provide some of the only evidence to date on the impacts of time limits. Furthermore, the policy variation across the three states nicely reflects the variation across state TANF plans.

Our choice to use experimental data and methods is not incidental. As discussed in Blank (2002) and formalized in Bitler, Gelbach & Hoynes (2003, (Papers and Proceedings)), identifying the impact of TANF using nonexperimental methods is difficult given that (*i*) TANF was implemented in all states within a very short period, and (*ii*) the implementation took place during the strongest economic expansion in decades. Since the relevance of many of our findings lies in their ability to detect heterogeneous treatment effects, we believe it is critical that our results not depend importantly on nuisance issues related to selection bias. To this end, experimental data provides a setting where identification is clear and essentially incontrovertible.

Our emphasis on heterogeneity is motivated partly on substantive grounds, as the focus on mean impacts has not gone unquestioned. For example, in a recent Joint Center for Poverty Research newsletter entitled "What Policymakers Want to Know," Cabrera & Evans (2000) ask "What is the variability of response to welfare reform among families?...Typically, research findings are reported in terms of the average response of the welfare reform population with respect to some behavior or status of interest. This focus diverts attention from the subgroup of families that might be struggling, even when most are not." Such concerns have led a few analysts to consider distributional concerns. Schoeni & Blank (2003 (Papers and Proceedings)), in arguably the most comprehensive paper examining distributional impacts, compare the full distribution of the income-to-needs ratio before and after TANF and find increases at all but the very lowest percentiles. However, as the authors note, their simple before-and-after methods cannot distinguish impacts of TANF from the effects of strong labor markets. With the exception of these results and those in a few other papers,[2] the most common approach to addressing distributional concerns is to estimate

[2]Schoeni & Blank (2000) compare the 20th and 50th percentiles of the CPS family income distribution before and after implementation of TANF. They find negative (but insignificant) impacts of TANF on the 20th percentile, and positive and significant impacts on the 50th percentile for a sample of women with less than a high school education. Some of the MDRC waiver evaluations (e.g., Bloom, Scrivener, Michalopoulos, Morris, Hendra, Adams-Ciardullo & Walter (2002) and Bloom, Kemple, Morris, Scrivener, Verma & Hendra (2000)) include estimates comparing the fraction of treatment and control group members with income in broadly defined categories. This approach, which is

mean impacts for subgroups of the population (defined using education, race, and welfare and employment history) thought to be particularly at risk for welfare dependence.[3] Michalopoulos & Schwartz (2000) review 20 randomized experiments, concluding that "Although the programs did not increase [mean] income for most subgroups they also did not decrease [mean] income for most subgroups" (p. ES-10). Grogger et al. (2002) summarize both nonexperimental and experimental evidence concerning mean impacts as follows: "the effects of reform do not generally appear to be concentrated among any particular group of recipients" (p. 231).

The focus in the literature on mean impacts contrasts with the very strong predictions that labor supply theory makes concerning welfare reform and heterogeneity. Consider, for example, Figure 1, which shows a stylized budget constraint in income-leisure space before and after Jobs First (whose characteristics we describe in detail below). Jobs First dramatically increased the disregard for calculating welfare benefits. In the pre-reform AFDC program, benefits were reduced dollar-for-dollar with increased earnings, leading to the horizontal portion of the budget set (which corresponds to a 100% tax rate). Under Jobs First, recipients retain their *entire* benefit payment (a 0% tax rate) for earnings up to the poverty line (unearned income was still taxed away). Labor supply theory makes clear predictions about this reform. Women who would choose to participate in welfare and not work when they face AFDC rules will increase earnings, provided their wages exceed a reservation level. By contrast, some women who would not participate under AFDC rules will decrease earnings (to the poverty line or below) to become eligible for Jobs First.

We illustrate these labor supply predictions in Section 2 by simulating effects of the Jobs First disregard reform using estimates from Moffitt's (1983) classic paper on welfare and labor supply. These results suggest that the disregard reform yields a proportionately large (though small in absolute terms) earnings increase among simulated women who would participate in welfare under AFDC rules. At the same time, the reform induces a significant number of women who would not participate under AFDC rules to reduce their earnings substantially in order to gain welfare eligibility under Jobs First rules.

Our empirical findings may be summarized with two broad conclusions. First, we find evidence

essentially a tabular form of histogram plots, is similar in spirit to the approach we take.

[3]For example, Schoeni & Blank (2000) find that welfare reforms led to increases (insignificant in the case of TANF) in mean family income for female dropouts in the CPS. Using similar data, Bennett, Lu & Song (2002) find that TANF is associated with reductions in the income-to-needs ratio for poor children who live with a single parent having less than a high school education.

of substantial heterogeneity in response to welfare reform. Second, the heterogeneity is broadly consistent with the predictions of labor supply theory. Figure 2 previews our empirical results and illustrates these findings clearly. The top panel plots deciles of the annualized earnings distribution among treatment and control group observations in the first two years following implementation of Jobs First (before time limits bind). The difference between these lines at a given decile is an estimate of the reform's treatment effect on earnings at that quantile.

We plot these QTEs in the bottom panel of Figure 2. The variation in the impact across the quantiles of the distributions is unmistakably significant, both statistically and substantively. As theory predicts, the earnings QTE is zero at the bottom two deciles, rises (to a peak of about $1,000) for the middle deciles, and then falls, becoming negative (about -$1,000) for the highest earners. Such results call into question the focus on mean impacts, as well as some of the substantive conclusions in the literature.[4]

This discussion raises the issue of entry effects. The apparent inability of experiments to measure entry effects has been widely noted. For example, consider the following quotation from Committee on Ways and Means (2000, p. 1443):

> Whether increases in work effort by those already on the rolls are canceled out by reduced work effort by those not previously eligible for benefits (as found in earlier nonexperimental research) cannot be told from these evaluations. The findings of these evaluations are limited to the population studied: those who already applied for and/or received cash assistance.

This view is correct when only mean impacts are available. However, as Figure 2 shows, QTE estimators allow us to observe reductions in earnings at higher quantiles of the earnings distribution.

[4]It is important to recognize that the QTE at quantile q does not necessarily represent the treatment effect for people whose quantile in the untreated distribution is q. This is because of the important distinction between the quantile treatment effect and the distribution across the population of treatment effects. This distinction is essentially that between marginal distributions and joint distributions. An assumption that guarantees that the QTE at q equals the treatment effect at q is *rank preservation*: a person's rank in the treated distribution is the same as her rank in the untreated distribution. When there is heterogeneity in both preferences (i.e, the slope of indifference curves over leisure and consumption) and parameters (i.e., wages and nonlabor income), this assumption is implausible if taken literally. However, when the QTE at the top of the distribution is non-positive, we can say for sure that treatment effects at the top are non-positive (since those at the top can only move to even lower quantiles if their ranks change). Similarly, non-negative quantile treatment effects at the bottom of the distribution necessarily imply non-negative treatment effects for those at the bottom of the untreated distribution. We defer a detailed discussion of these and related methodological issues until section 4. For now the key fact is that heterogeneity in the QTE implies heterogeneity in the distribution of treatment effects.

This finding, which is predicted by labor supply theory, would not be possible using only mean impacts.

While not widely considered in the literature, heterogeneous impacts of welfare reform are thus clearly predicted by labor supply theory and evident in our data. These facts provide the essential motivation for our analysis, and they raise an important methodological question: could these results be revealed with mean impact analysis using subgroups? In at least some important cases, we find striking evidence that the answer is a resounding *no*. We estimate causal impacts of welfare reform on the earnings, transfer, and income distributions not only for the full sample (as above), but within subgroups used in previous studies. We find that *intra-group* variation in treatment effects often exceeds the *inter-group* variation. Importantly, under plausible assumptions our results suggest the possibility that welfare reforms reduced income for a substantial fraction of treatment group members, especially after time limits take effect.

The remainder of the paper is organized as follows. In section 2, we discuss expected effects of these waivers on earnings, transfers, and income, and we provide simulation-based evidence to demonstrate the likely importance of heterogeneous treatment effects. In section 3, we present the main results, for Connecticut's Jobs First. In section 4, we briefly discuss a standard model of potential outcomes and quantile treatment effects. We then discuss some important methodological issues concerning the difference between quantile treatment effects and distributions of treatment effects. Section 5, to be added in a future version, will present results for FTP and MFIP. Section 6 concludes.

2 Theory of Labor Supply and Heterogeneity

In this section we use standard labor supply theory to discuss the expected effects of Jobs First and similar reform programs. We begin in subsection 2.1 by considering only the effects of the expanded disregard associated with Jobs First. The same basic arguments apply when considering the less-generous disregard expansions in FTP and MFIP. In subsection 2.2, we present some illustrative results from a simulation of labor supply and welfare participation behavior using Moffitt's (1983) estimates. Lastly, we discuss the expected labor supply and participation effects of the other major components of reform in subsection 2.3.

2.1 Labor supply theory and expanded disregards

We start by ignoring the existence of the welfare system and any other taxes and transfers, so that a woman faces constant wage and nonlabor income values, w and N. Under standard assumptions, we can write the choice of labor hours that satisfies the first-order condition for utility maximization as a function of wages and nonlabor income, say $h^*(w, N)$. Assuming the second-order condition for optimality is satisfied, a woman will choose to work positive hours for pay if and only if $h^* > 0$.[5] Replacing the inequality with an equals sign and holding constant nonlabor income, this relationship defines a reservation wage, $w^*(N)$, such that the woman works positive hours if and only if $w > w^*(N)$. In terms of Figure 1, this reservation wage is the one for which the woman's budget line is exactly tangent to her indifference curve at 0 hours of labor supply.

Now we introduce a stylized AFDC system, which may be characterized as providing welfare benefits equal to $B_a = \max[0, G - wh^*(0, G) - N]$, where G is the maximum welfare benefit. Optimal hours are evaluated at a wage of 0 because of the implicit 100% tax rate, and at the nonlabor income value of G. The max operator is used because benefits cannot be negative. In the absence of any welfare stigma, utility is a function only of hours of labor (usually a bad) and consumption, which equals the maximum benefit if the woman participates in welfare. We may write maximized (indirect) utility if the woman participates in welfare as $U^*_{1a} = U(h^*(0, G), G)$. If she does not participate, maximized utility is $U^*_0 = U(h^*(w, N), N + wh^*(w, N))$. Because of the widely observed phenomenon of eligible nonparticipation, we add the possibility of so-called "flat stigma," which may be thought of broadly as any factor that lowers the likelihood of welfare participation, all else equal. Denoting this flat stigma ϕ, a woman participates in welfare when facing AFDC rules if and only if $U^*_{1a} - \phi > U^*_0$. While we use Moffitt's positive estimate of ϕ in the simulations below, the remainder of this discussion assumes for simplicity that $\phi = 0$.

As is well known, if work hours are a bad, women will always participate in welfare and work 0 hours when $h^*(w, N) \leq G$, so that $B_a = G$ for women who participate. This is because of the implicit 100% tax rate on AFDC: there is no point working G/w (or any other positive amount of) hours when one can work 0 hours and have the same consumption. Moreover, the 100%

[5] We mention second-order conditions because the preferences that generate a linear desired-hours function, which are used in Moffitt (1983), can yield somewhat unconventionally shaped indifference curves. These issues are discussed in detail in Appendix A, which considers the simulation in detail.

tax rate on AFDC introduces a nonconvexity in the budget set. There is thus some nondegenerate neighborhood of earnings values above G such that women whose optimal hours would yield earnings in that neighborhood will also participate and not work. In other words, the indifference curve through the "corner" point $(0, G)$ is strictly higher than the indifference curve through (h, G) for all $h \in (0, G/w + \epsilon]$, some $\epsilon > 0$. This discussion implies that the existence of AFDC discourages women who would be income-eligible in the absence of the program from working, while encouraging some women who would not be income-eligible to cease work altogether, instead participating in welfare.

To analyze the impact of Jobs First's more generous disregard policy, note that benefits under Jobs First may be written as $B_j = G \times 1(wh^* \leq FPL)$, where FPL is the federal poverty line and $1(\cdot)$ is the indicator function equal to 1 when its argument is true and 0 otherwise. The net wage under Jobs First is always the same as the gross wage, while nonlabor income equals $G + N$. Maximized utility on Jobs First is $U^*_{1j} = U(h^*(w, G+N), G+N)$, while maximized utility off of Jobs First is the same as before. Since the budget set on welfare has been expanded, utility on welfare cannot fall due to the Jobs First disregard. It follows that no one will leave welfare because of the expanded earnings disregard, while some women will enter.

Consider first women who would locate at the corner of the AFDC budget set, working 0 hours and receiving the maximum benefit. For these women, the effective wage rises from 0 to w, and nothing else changes. For women whose indifference curves at the corner are more steeply sloped than the budget line with slope $-w$, the expanded disregard will have no impact, since it leaves their effective wage below their reservation wage. However, women whose indifference curves at the corner have slope between $-w$ and 0 will enter the labor force because of the expanded disregard. For these women, the disregard's impact on hours worked and earnings will be positive; since the tax rate is 0 while on Jobs First, these women also will not lose any benefits.

Now consider women who would not participate in welfare under AFDC. For these women, the indifference curve tangent to the slope $-w$ budget line is higher than any indifference curve on which the woman participates in AFDC. The earnings disregard adds a new, much higher kink (or elsewhere on the slope $-w$ budget line through it) to the budget set where earnings equal the federal poverty line and benefits equal the maximum benefit. For women whose indifference curve at this new kink is higher than the one tangent to the slope $-w$ budget line, the expanded disregard induces new welfare participation.

To see this, consider first those new participants whose earnings when subject to (but not participating in) AFDC would be below the federal poverty line. For them, the participation effect may be thought of as mechanical; earnings are less than the new eligibility cutoff, so they are income-eligible without any change in behavior. Barring large stigma effects, these women will enter welfare. The will also reduce work hours, since their effective wage on Jobs First is unchanged but their nonlabor income has risen substantially. Next consider AFDC non-participants whose earnings off welfare exceed the federal poverty line. For some of these women, it will be optimal to reduce earnings below the federal poverty line in order to gain Jobs First eligibility: lost earnings will be offset (at least partly, and possibly more than fully) by welfare income. This participation effect may be thought of as behavioral. Both the mechanical and behavioral participation effects are associated with reduced earnings and work hours, and an increase in welfare income.

2.2 Simulated effects of the expanded disregard

To illustrate these predictions, we generate 2000 simulated observations having demographic characteristics equal to the means of Moffitt's (1983) sample and a lognormal wage distribution chosen to approximately match the mean wage reported by Moffitt. We use Moffitt's parameter estimates to predict each simulated woman's optimal hours of labor supply on and off welfare under AFDC as well as her optimal welfare participation decision. We then repeat the simulation of optimal decisions under the Jobs First disregard. Details of the simulation are presented in Appendix A.

Table 1 reports key results from this simulation. The first column shows mean outcomes under the AFDC rules for *AFDC Participants*, those women who would participate in welfare under AFDC rules. The second column reports mean impacts on these women of implementing the Jobs First expanded disregard; thus the sum of the first and second column yields mean outcomes under Jobs First. The results show that AFDC participants work very few hours—only 2.05 per week, and that this amount is doubled by implementing the expanded disregard.[6] Earnings, expressed in units of the 1975 minimum wage, rise similarly.[7] Total income rises by nearly half, as earnings and

[6]One might wonder why AFDC participants would ever work positive hours. As we discuss in Appendix A, the linear desired-hours function used by Moffitt implies that indifference curves can be U-shaped (or even inverted-U-shaped). With such preferences the marginal disutility of work hours is sometimes zero, so that positive work hours can be optimal with a net wage of 0.

[7]The 1975 minimum wage was \$2.10 in nominal dollars. The maximum combined AFDC-plus-Food Stamps weekly

welfare income both rise (welfare income rises because women who work and lose benefits under AFDC's 100% tax rate no longer lose those benefits under Jobs First). The women in these two columns constitute 24.6% of the simulated population; thus the AFDC participation rate is 24.6% (in Moffitt's sample, 35% participate).

The last four columns of the table concern *AFDC Non-participants*, women who would not participate in welfare under AFDC rules. The first two of these columns report results for all such women. The results show that 5.8% of AFDC non-participants participate in Jobs First, so that the overall treatment effect on welfare participation due to the expanded disregard is $0.754 \times .058 = 0.04$. As predicted, hours fall, by about 5%; earnings fall by about the same amount. Total income also falls, as lost earnings more than offset increased welfare income.

The final two columns of the table focus only on the 4.3% of the population who are "switchers": simulated women who do not take welfare under AFDC rules but do under the Jobs First expanded disregard. Under AFDC rules, these simulated women work substantial hours—more than 48—and the expanded disregard causes them to reduce work effort by more than half. The result is a similar loss in earnings, which more than offsets their receipt of the maximum welfare benefit.[8]

These simulation results illustrate that we can expect the population of women who would participate in welfare under AFDC to increase work effort and earnings under an expanded disregard, as has been widely discussed. However, the results illustrate that for a small but nontrivial group of women, the very same policy change can dramatically lower labor supply and earnings. Simulation of a simple and standard economic model yields considerable heterogeneity in the effects of the expanded disregard. This compositional heterogeneity would be unobservable using mean impacts.

---

benefit in Connecticut for a family of three was $45.13, or 21.49 minimum wage units.

[8]It might be surprising that among AFDC non-participants, Jobs First participants have greater hours under AFDC rules than do Jobs First non-participants. This occurs because of the assumed independence of the wage and the stochastic part of stigma; this independence is discussed in Appendix A. Any woman with both low stigma and a low wage will participate in AFDC, while women with both high stigma and a high wage will participate in neither AFDC nor Jobs First. The "switchers" are women who have moderate values of both parameters or a high value of one and a low value of the other. As expected, experimentation with the simulated observations suggests that the comparatively high earnings of switchers are due to a large number of low-wage, high-stigma women among the women who participate in neither program.

2.3 Expected effects of the other reform characteristics

Besides disregards, the other principal reform measures are binding minimum work requirements and time limits. Work requirements may be modeled as removing the portion of the budget set in which work hours are less than the minimum and welfare benefits are received. This change has two straightforward implications. First, it will cause some Jobs First participants to increase work hours in order to maintain eligibility. Second, it will cause some other women to forego welfare participation under Jobs First in order to avoid working more hours on welfare than is optimal. In principle, it would be straightforward to incorporate this reform in the simulation model.

Effects of time limits would require a dynamic model, or at least a repeated model, rather than the one-period simulation model discussed in the previous subsection. However, the one-period model does allow some general intuition. Note first that effects of time limits may be classified as mechanical and behavioral. The mechanical effects are most easily seen by assuming women are perfectly myopic, so that each period they make decisions as if these had no impact on future budget sets. Myopic women who participate in welfare enough periods to hit the time limit will eventually lose eligibility for Jobs First. When that happens, labor supply and earnings will unambiguously increase if leisure is a normal good (as implied by Moffitt's estimates), because lost welfare benefits will induce more hours of work. The combination of lost benefits and increased earnings means the effect on total income is ambiguous. Behavioral effects of time limits can occur when women are forward-looking. Knowing that each period of welfare participation reduces the remaining stock of eligibility, forward-looking women will choose not to participate even when eligible unless their situation is sufficiently dire; this behavioral change may be thought of as a progressive increase in the stigma parameter ϕ.[9]

Since the behavioral effect reduces effective nonlabor income (and the effective wage is the same), it must increase labor supply and earnings, while reducing welfare income. Thus the mechanical and behavioral effects of time limits affect labor supply and income similarly. Relative to the expanded earnings disregard only, a program with time limits may either increase or decrease a woman's total income. Lastly, if time limits concern only the number of periods a woman participates, rather

[9]Using variation in the remaining stock of benefits due to the interaction of time limit rules and children's ages, Grogger (Forthcoming) and Grogger & Michalopoulos (Forthcoming) provide considerable evidence of some such behavioral responses.

than the amount of assistance conditional on participation, then they will have no effect on labor supply and earnings in periods when women are on welfare under the expanded disregard.

3 Jobs First impacts on Earnings, Transfers, and Income

In this section we summarize the impacts of Connecticut's Jobs First program on the earnings, transfer and income distribution. Results for FTP and MFIP will be added in a future version and presented in Section 5. Our analysis is based on public-use data provided by the Manpower Demonstration and Research Corporation (MDRC), which conducted the contract evaluations of the three waiver programs. While many states sought and obtained waivers during the 1990s (Committee on Ways and Means (2000)), we use these three states for four reasons. First, under current law, states are not required to evaluate their TANF programs and "most of the information available on the impacts of welfare reform initiatives has been provided through waiver evaluations begun under AFDC" (Committee on Ways and Means (2000, p. 1475)). Second, these programs are among the few waiver programs containing the essential features of the subsequently established TANF requirements. In particular, Jobs First and FTP have time limits, strong work requirements, financial sanctions, and enhanced work incentives. Except for time limits, MFIP also has these provisions. With few changes, these programs were adopted as the states' TANF programs.[10] Third, Jobs First (because of its short time limit), and FTP (because it implemented its time limit early), are often cited as providing the best available evidence on the impact of time limits (e.g., Blank (2002), Grogger et al. (2002), and Grogger & Michalopoulos (Forthcoming)). Last, the three states exhibit important variation in their program rules, similar to the cross-state variation observed in TANF plans.

3.1 The Jobs First Reform, Data, and Mean Impacts

Table 2 summarizes the main features of Jobs First, as reported in MDRC's report (Bloom et al. (2002)). The table also includes a summary of the pre-existing AFDC program for comparison.

The Jobs First waiver contained each of the main elements in the eventual TANF legislation

[10] HHS is funding the continued study of key waiver states to examine TANF's impacts. Nine states have been chosen for a broad impact analysis, and a smaller set of five states has been chosen for a child impact analysis. Our states are included in both studies.

including time limits, work requirements, financial sanctions, and changes to earnings disregards. Jobs First's 21-month time limit is currently the shortest in the U.S. (Committee on Ways and Means (2000)). About 29% of the treatment group reached the time limit in the first 21 months of the evaluation period, and more than half reached the time limit within four years after random assignment (Bloom et al. (2002)).[11] TANF work requirements include both "work-first" models that focus on immediate employment and "mixed" programs assigning some participants to work-first programs and others (typically the least job ready) to education and training activities. Jobs First operated a work-first program, exempting only those with a child younger than one. Under Jobs First, recipients who fail to comply with work requirements face financial sanctions that reduce their grants by 20–100% for three months, depending on the number of prior instances of noncompliance.[12]

As discussed above, Jobs First (and the majority of state TANF programs) complement the time limit, work requirement, and sanction "sticks" with some important "carrots," namely increased earnings disregards. The general AFDC monthly benefit formula is $B_A \equiv G - t_{Am} \max[0, (E - D_{Am})]$, where B_A is the net benefit, G is the maximum benefit, t_{Am} is the benefit reduction rate when the AFDC spell is in month m, D_{Am} is the flat disregard in spell month m, and E is earnings. Under pre-existing AFDC rules, the flat disregard is $D_{Am} = 120$ for $m \leq 12$ and $D_{Am} = 90$ for $m > 12$, while the tax rate is $t_{Am} = 2/3$ for $m \leq 4$ and $t_{Am} = 1$ for $m > 4$. Thus the long run implicit tax rate on earnings above the flat disregard is 100%.[13] By contrast, Jobs First has a very generous disregard policy. Its formula is given by $B_J \equiv 1(E < FPL) \times G$, where FPL is the federal poverty line and $1(\cdot)$ is the indicator function. Thus, Jobs First disregards *all* earned income up to the poverty line, implying a tax rate of 0 on earnings up to the poverty line. As Figure 1 illustrates, there is a large notch at the poverty line, where welfare benefits are completely lost. In practice,

[11] The stringency of the time limit was tempered by the use of six-month extensions granted to participants who were complying with program requirements but had earnings below the maximum grant.

[12] Monitoring was poor, and only 8–13% of the treated were sanctioned, compared to 5% in the control group (Bloom et al. (2002)). To evaluate fully the stringency of the sanctions, one would need to know the rate of noncompliance. We do not know of any such data.

[13] In addition to earnings, most non-labor income (e.g., child support after a $50 pass-through) was implicitly taxed at the rate t_{Am} under AFDC. However, after 19XX, all income from the Earned Income Tax Credit (EITC) was disregarded by AFDC. Therefore, the effective tax rate on earnings is one minus the EITC subsidy rate, which was 0.402 for a woman with two or more children. Moffitt (1983) argues that the effective marginal tax rate on AFDC is substantially lower than 100% due to deductions (e.g., for child care and work expenses). Some of these expenditures (e.g., transportation) may have no *per se* value as consumption goods, so deductions for them are appropriately ignored. Others (e.g., work clothing) would need to be accounted for in a structural model.

these policies interact with the EITC, so that the effective hourly wage for a Jobs First recipient with earnings in the EITC phase-in range equals 140% of the hourly wage, compared to just 40% under the AFDC program.[14] In addition to these changes, Jobs First increased asset limits and implemented a partial family cap, as noted in the "other policies" section of the table. It is also important to note that Connecticut has a relatively generous benefit level of $543 for a family of 3.

The experimental design for Jobs First selected samples randomly from the stock of existing recipients (the recipient sample) and the flow of new recipients (the applicant sample) between January 1996 and February 1997 and had a follow-up period of four years.[15] Each sample member is then randomly assigned to either the treatment group (waiver program) or the control group (AFDC program). The public-use data file has information on 4,803 single parents (2,396 treatment and 2,407 control), all of whom were served by either the New Haven or Manchester welfare offices.[16]

The primary data sources for the public-use files are administrative. Monthly welfare participation and benefit data are taken from state welfare files, while quarterly earnings data come from employer reporting to the unemployment insurance (UI) system. While separate variables are provided for welfare and Food Stamps, we sum these variables to create a single variable that we call total transfer income. Earnings and transfer data are provided for eight quarters prior to random assignment and for up to 16 quarters following random assignment. Our total income variable is the sum of transfer benefits and earnings and is measured quarterly.[17] Demographic data—including information on educational attainment, age, race and ethnicity, marital status, and work history of the sample member—are collected at an interview prior to random assignment.[18]

[14]The expanded disregard also affects Food Stamp eligibility and benefits. Under AFDC rules, eligibility for AFDC conferred categorical eligibility for Food Stamps. Increasing the earnings disregard will, in general, lead to an increase in eligibility for welfare and an increase in Food Stamp eligibility. However, losing eligibility for welfare benefits (e.g., through time limits) need not eliminate Food Stamp eligibility, since one could still satisfy the Food Stamps need standards. In addition, Jobs First Food Stamps rules mirrored cash assistance rules, with Food Stamps benefits determined after disregarding all earnings up to the poverty line.

[15]Sampling from the stock of existing recipients takes place over the same time period as the applicant sample. Recipients are required to visit the welfare office for eligibility redetermination at regular intervals, at which point recipient sampling took place.

[16]Two-parent families and child-only cases were (along with single-parent families) randomized into the Jobs First (and FTP) experiments. However, since neither the MDRC reports nor the public-use files include data for these groups, our analysis is limited to single-parent families. The MFIP public-use file does include a two-parent sample, but for comparability we also limit our analysis of MFIP to single-parents cases.

[17]To preserve confidentiality, MDRC rounded several key variables before releasing the public-use data. For example, the quarterly earnings data is rounded to the nearest $100, and the AFDC and Food Stamps payments are rounded to the nearest $50.

[18]In each of the three experiments, MDRC conducted a survey on a subset of the sample about three years after

This experimental design provides an excellent setting for our analysis of the impacts of welfare reform on the income distribution, though there are some drawbacks. First, administrative data undercount earnings and do not measure the income of other family members, missing potentially important sources of family income.[19] Second, as mentioned above, sampling only from women who receive welfare at a fixed time precludes measuring impacts of reform on changes in entry into welfare among women not included in the study sample. Grogger, Haider & Klerman (2003 (Papers and Proceedings)) have shown that reductions in entry rates into welfare are responsible for a large fraction of the fall in welfare caseloads in the 1990s. While this is an important issue, the population we analyze is nevertheless well-defined, important, and has been extensively analyzed in the literature. Moreover, effects of Jobs First on later re-entry behavior among sample participants is observable. In any case, in the present context we strongly feel that the gains to using experimental data—in terms of the ability to identify true causal impacts—are large compared to the limitations inherent in the analysis.

Experiments also sometimes experience contamination of the control group by the treatment. This issue may be of particular concern here because these waivers were taking place during a time of significant change in welfare programs at the state and federal levels. This could cause control group members to misunderstand the rules they face, or to anticipate rules they might face in the future. In Connecticut, Jobs First was implemented statewide, and only those in the evaluation's control group remained covered by AFDC. About 23% of control group members surveyed by MDRC reported that they faced time limits, compared to 90% in the Jobs First treatment group (Bloom et al. (2002)).[20]

We report summary statistics for the Jobs First experimental sample in Table 4. For comparison

random assignment. These data, which we do not use, have been used in various studies to analyze impacts on other measures of family and child well-being.

[19] UI data excludes earnings from self-employment, from employment by the federal and some state and local governments, and from some agricultural employment. Further, state administrative files capture only those earnings and welfare benefits received within the state; these files report zeroes for all people who move out of state.

[20] Contamination is also cited as a problem in Florida, where the state's TANF program (Work and Gain Economic Self Sufficiency, or WAGES) was implemented statewide in October 1996, in the middle of the FTP evaluation period. Control group recipients were told, however, that they were not subject to the WAGES program for the duration of the evaluation. Still, it is cited as being a possible factor in the significant reduction in treatment effects in the last year of the follow-up period (Bloom et al. (2000)). This issue seems to be less of a concern in Minnesota, where TANF was not implemented until the last six months of the MFIP evaluation (Miller, Knox, Gennetian, Dodoo, Hunter & Redcross (2000)).

purposes, we also provide summary statistics for the national AFDC caseload.[21] The Jobs First sample generally mirrors the characteristics of the national sample, although it does have higher rates of being never married, Hispanic, and less-educated among recipients compared to the national caseload averages.[22] The baseline characteristics between the treatment and control groups are very close, reflecting the random assignment process. The two exceptions are that the fraction with prior employment is lower in the treatment group, while the fraction with recent prior welfare use is greater, when compared to the control group.

In addition to analyzing the full sample for each program, we investigate differences across groups and present results for the most- and least-disadvantaged subgroups of the caseload, as defined by MDRC. The most-disadvantaged group has education less than a high school degree or GED, has been on welfare for at least 22 of the 24 months prior to random assignment, and has no earnings in the year prior to random assignment. The least-disadvantaged group has none of these three characteristics. Table 4 also presents the mean characteristics for these subgroups.

Table 5 presents mean impacts for Jobs First. We present results for the full sample and for the most- and least-disadvantaged subgroups. For each sample, we present both unconditional and regression-adjusted treatment-control differences. In light of Jobs First's 21-month time limit, we provide treatment effects separately for years 1–2 and years 3–4.[23] The traditional reason to add covariates with random assignment is to increase precision of the estimated treatment effects. However, as noted in Bloom et al. (2002) and shown in Table 4, the Jobs First control group by chance had greater labor force attachment and lower welfare use prior to random assignment than did the treatment group. Because of this correlation, regression-adjusted mean impacts for Jobs First are statistically different from the unconditional mean impacts for two outcome variables: earnings and transfer income in the first two years.[24]

[21] The estimates for the national caseload are constructed with data from the March 1995 CPS. The sample includes all women aged 16–54 who have an own child in the household and whose family was reported to receive positive AFDC income in the prior calendar year.

[22] The table also shows that Jobs First sample members are more likely than the CPS sample members to have worked in the preceding year. This is most likely an artifact of differences in the sample selection. The administrative data measure employment in the 12 months prior to random assignment. Our CPS sample contains women who report receiving AFDC income at any point during calendar year 1994. Thus we do not have a full 12-month period prior to the month of AFDC receipt, which will cause the AFDC sample to have less employment relative to the experimental sample.

[23] Because of recoding in the public-use files, the estimates do not replicate exactly the results in the MDRC reports.

[24] To estimate regression-adjusted impacts, we use the following baseline characteristics used in the MDRC evaluations: dummies for employment, and earnings and AFDC income levels for the four quarters preceding random

The estimates for the years 1–2 show that Jobs First led to large income gains in the first two years for all groups. Interestingly, the impact on income is fairly similar for the least- and most-disadvantaged groups. However, as noted in Bloom et al. (2002), the composition of income changes is very different for the least- versus the most-disadvantaged groups. Income gains for the most-disadvantaged recipients come almost entirely from increased earnings, while gains for the least-disadvantaged group are due to increases in transfer income. This result can neither refute nor confirm the theoretical predictions, since it is possible there is significant within-group heterogeneity, but it is consistent with expectations: the most-disadvantaged women can be expected to receive large quantities of welfare income and are encouraged to increase work. The least-disadvantaged can be expected to have relatively high earnings and encouraged to reduce earnings and increase welfare participation. In years 3 and 4, the significant income gains disappear. The mean increase in earnings does not change much over the period, but in the later years it is offset by an equal- (or greater-) sized reduction in transfer income.

3.2 QTE Results for Full Sample

In this section we present estimates of the impact of welfare reform on the distribution of income. Our analysis simply compares treatment and control groups and does not control for any additional variables.[25] For all variables we examine, we average values across years to create annualized values of outcome measures.[26]

---

assignment. The systematic changes in mean impacts raise concerns about computing QTE differences using simple differences in the marginal distributions. The direct analog to regression-adjusted differences, which are computed as the coefficient on the treatment dummy when linear terms in the covariates are included, is to estimate a quantile regression with the same linear terms included. However, this procedure would constrain the QTE given each vector of covariates to equal the QTE given any other vector, which imposes potentially important restrictions on the heterogeneity allowed for treatment effects. To address this concern, we use Firpo's (2003) recently developed semiparametric QTE estimator. Essentially, this estimator involves estimating the simple QTE using the propensity score for the treatment dummy to weight the regression. Firpo (2003) shows that the true QTE is identified by this procedure as long as the propensity score is consistently estimated in a first-stage. Initial results using this estimator with bootstrapped standard errors suggest that all of our qualitative conclusions are robust to controlling for covariates in this fashion.

[25] See footnote 24 on the preceding page for a discussion of covariate-robust estimates.

[26] Annualized values provide a reasonable basis to assess the impact of reform on long run outcomes. However, there are two reasons one might not want to annualize. First, if women are unable to significantly smooth consumption across sub-annual periods (as assumed in Grogger & Michalopoulos (Forthcoming)), then outcomes in higher-frequency periods would be of substantive interest. Second, if there is significant volatility in any one individual's outcomes, annualizing over a two-year period may compress the estimated QTEs toward the mean impacts, obscuring the greater variation one would observe in any given time period. *De jure* welfare eligibility and payment amounts are based on monthly earnings, though MDRC's implementation report suggests that in practice quarterly

Figure 3 plots QTE estimates for earnings, transfers, and total income in the two-year periods following implementation of Jobs First. The solid lines are the estimated QTEs, while the dotted lines provide 90% confidence bounds based on a two-tailed test. For reference, the mean impact (and its 90% confidence interval) is shown by the horizontal solid and dotted lines.[27] Panel (a) repeats the bottom panel of figure 2 and shows that Jobs First led to increases in the earnings distribution in years 1–2 for deciles 3-8 on the order of $250 to $750. Increases in the earnings quantiles are more uniform than those for income, with the exception of the large reduction of more than $1000 at the highest decile. This pattern of earnings impacts is strikingly consistent with the incentives outlined in section 2 above. Earnings at the bottom deciles are unchanged, suggesting that lowest-earners may remain on welfare and not work. Women who are more likely to have wages above the reservation levels are expected to respond to the large increase in the net wage and increase earnings, and the QTEs for the middle deciles of the earnings distribution are indeed positive. Lastly, some high-earning recipients are expected to reduce hours and earnings, which is consistent with the reduction in earnings at the top decile.

Panel (b) shows the extremely large positive impacts of the program on the distribution of transfer income, reflecting Jobs First's generous disregard. The QTE is above $500 at all quantiles, with peak impacts of about $1500, or nearly twice the mean impact, at the third decile. Again assuming rank preservation, the smallest gains go to the largest welfare users, because they are more likely to be receiving welfare in the counterfactual no-reform state. By contrast, the largest gains occur for the lower half of the distribution—representing the group that is encouraged by the enhanced disregards to stay on aid.[28]

earnings are used for eligibility and benefit determination. To the extent that offered wages vary across quarters, using quarterly values would be a better way to test the theoretical predictions. One would expect a greater range in QTEs computed using quarterly values as the unit of analysis, because higher-frequency variation in outcomes will be smoothed by using averages of individual outcomes over a two-year period. The question of which quarters to use then arises, and an additional problem is that the sampling variance of estimated treatment effects will likely be greater using only one quarter. A solution to both of these problems is to compute the QTE separately for each quarter and quantile, and then average the quarterly estimates of the QTE at each quantile q. Unless there is no cross-quarter, within-individual variation in the outcome variable, this procedure will lead to lower sampling variance than using any given quarter, and it also allows consistent estimation of the greater range of QTEs. Preliminary investigation confirms that estimates computed this way are relatively precise and have a greater range than those reported below. Estimating the sampling variance of these estimates requires extensive bootstrapping and is thus very time-consuming; we will add these results to a future version of the paper.

[27]The accompanying figures for treatment and control distributions for each of these outcomes are provided as Appendix Figures 1–3.

[28]It is important to note that the q^{th} quantile of the earnings distribution will not—and empirically does not—generally correspond to the q^{th} quantile of the transfer distribution. One would typically expect these quantiles to

Panel (c) concerns total income (earnings plus transfers) for the first two years of the Jobs First program. The QTE for total income is everywhere positive and increasing in decile, ranging from less than \$100 at deciles 1 and 2 to more than \$2000 for deciles 7 and above. If we assume rank preservation in the income distribution, this implies that the largest gains went to the best-off recipients. Alternatively, with some degree of rank reversal, some of the best-off women face large income losses, while other women achieve large income gains.

Jobs First's impact on the earnings distribution after time limits hit is similar in shape to the earnings impact for the first two years, but with a larger range. By contrast, the impacts on transfers and total income change dramatically with time limits. The impact of reform on transfer income in the second two years is eye-opening: welfare income declines substantially in the upper half of the distribution. Rank preservation seems particularly reasonable for the high-welfare use population more than two years after implementation of Jobs First. It is well-known that under AFDC, relatively few women have welfare spells exceeding two years, but that these women typically stay on aid for extended periods of time (e.g., Bane & Ellwood (1994)). Under rank preservation, the QTE results indicate that the most intensive welfare users lose substantial welfare income when time limits hit. By contrast, those in the lower deciles are essentially unaffected because they are low users in the absence of reform. Impacts on the distribution of total income are much lower and less variable—with a low of -\$500 at decile 3 and a high of \$500 at decile 6.

These tremendous differences in the first and second two-year periods, and the role played by the welfare rules, are explored further in Figure 4. This figure plots nonparametric (lowess) regressions of transfer income on earnings, with separate series for the treatment and control groups. The top panel, for years 1–2, shows that transfers are essentially constant as a function of earnings for the treatment group over a wide domain of earnings. By contrast, transfers decline steeply with earnings for the control group, reflecting the high implicit AFDC tax rate. This graph illustrates that even with no change in earnings (i.e., no behavioral response), treatment group members with positive earnings will receive increased transfers via Jobs First's generous earnings disregards. However, the bottom panel shows that in years 3–4, these differences disappear, as time limits hit and benefits discontinue.

be negatively associated, and generally they are.

3.3 Results for Subgroups

The main approach in the literature for measuring heterogeneity is to examine mean impacts within subgroups. To explore the importance of inter-group versus intra-group heterogeneity, we present the same figures for the MDRC-defined least-disadvantaged group (Figure 5) and most-disadvantaged group (Figure 6) in the Jobs First sample. As Table 5 and the horizontal lines in the figures show, examining mean impacts within these groups yields virtually identical impacts: an increase of about $1000 in years 1–2, followed by essentially a zero impact in years 3–4. The QTE impacts across the two groups, however, tell a completely different story. The least-disadvantaged group shows everywhere positive and large (up to $2,500 per year) changes in the income distribution in years 1–2. The most-disadvantaged group has *no* impact on the first five deciles, followed by large positive increases for the highest four deciles. Under rank preservation, this shows tremendous within-group variation in treatment effects; alternative joint distributions that allow for rank reversals would show even greater heterogeneity.[29] The impact on the highest decile in the most-disadvantaged group is almost $3,000, or three times the mean impact.

This variation is also evident for the impacts on the earnings and transfer distributions. For the least-disadvantaged group, the impact of Jobs First on the earnings distribution shows even more dramatically how the program leads to heterogeneous impacts due to basic labor supply incentives. Among this relatively skilled group, overall earnings do not increase substantially. At moderate earning levels, the Jobs First incentives are to increase earnings, while at higher levels the incentives are to reduce earnings and to gain welfare eligibility. Women at the very highest earnings levels are expected to stay off of welfare regardless of Jobs First, so the reform is expected to have no impact on earnings at the very top of the least-disadvantaged distribution.

The QTE estimates for the least-disadvantaged are strikingly consistent with these predictions. The reform increases transfer income at all deciles of the least-disadvantaged distribution in the first two years. It is helpful to remember that everyone in the sample is on welfare at the time of random assignment. The positive treatment effects at low deciles suggest that women who would be counterfactually low welfare users either respond to incentives to stay on aid longer or have their

[29] Using variance as the measure of heterogeneity, this fact is implied by the discussion of Frechet-Hoeffding bounds for superadditive functions in Heckman, Smith & Clements (1997, p. 498). The essential fact is that for the treatment effect distribution, the rank-preserving joint distribution yields minimum variance in the set of all joint distributions, given any two fixed marginal distributions.

benefits reduced less under Jobs First as their earnings rise. The transfer income increases are very large for the middle deciles—about $2,000. By contrast, for the least-disadvantaged Jobs First has zero impact on transfer income in the second two years after time limits hit. This occurs because even under AFDC rules, many less-disadvantaged women can be expected to leave welfare after two years have passed, and the generous disregard is no longer available to those of them who have stayed on welfare for the entire post-random assignment period.

The results for the most-disadvantaged show opposite patterns. For the highest four deciles, Jobs First leads to large increases ($2,000–$3,000) in earnings throughout the four-year period and zero impact on the lowest three deciles. This latter finding is extremely important in assessing whether a program like Jobs First—which combines the strongest reform stick and the strongest reform carrot currently in use in the nation—can be expected to yield self-sufficiency among the worst-off cases. The answer appears to be no. Even in the presence of a very generous disregard, and even after time limits may take effect, the impact on earnings for a significant fraction of the most-disadvantaged women is *zero*. While some of these women are likely exempted from time limits, these findings suggest little reason for confidence their earnings will later rise.

The transfer income results for the most-disadvantaged are equally striking. Transfers decline modestly at deciles 3–9 in years 1–2, and substantially at nearly all deciles in years 3–4. That is because this is a welfare-dependent group, one that would receive large amounts of cash assistance in the absence of reform. For the most disadvantaged, the point estimates imply that income under Jobs First is lower for more than half the distribution in years 1–2, and almost three-quarters in years 3–4. For the most part the 90% confidence intervals for these estimates contain 0. Unless there are substantial rank reversals, especially among women who start at the bottom of the most-disadvantaged income distribution, these results imply that income gains are largely concentrated among a relatively small, better-off group. If there are significant rank reversals involving women at the bottom of the distribution, then it follows that the women who wind up at the bottom must have lost considerable income.

4 Interpreting Estimated QTE

In this section we consider some important methodological issues. We begin in subsection 4.1 by briefly introducing the standard potential outcomes model that undergirds the preceding empirical results. In subsection 4.2 we discuss the measurement of QTE and their close cousin, treatment effects on distributions. We then turn in subsection 4.3 to the more difficult issue of measuring the distribution of treatment effects. Finally, in subsection 4.4 we discuss arguments for the relevance of both of these measures.

4.1 Potential outcomes

We give a brief discussion of a standard potential outcomes model.[30] Let the outcome of interest be Y_i, with Y_{i1} being i's outcome under the counterfactual state of receiving some treatment; similarly, Y_{i0} is i's outcome under the counterfactual state of not being treated. Let the treatment status for person i be indicated by the dummy D_i ($= 1$ if treated and 0 if not). We define the treatment effect for person i as $\Delta_i \equiv Y_{i1} - Y_{i0}$. We can write $Y_i = Y_{i0} + \Delta_i D_i$, so that $E[Y|D] = E[Y_0|D] + E[\Delta|D]D$, which suggests a standard regression model of the form $Y = Y_0 + E[\Delta|D = 1]D + u$, with the residual term satisfying $E[u|D] = 0$, $D \in \{0, 1\}$. This is the usual model implicit in studies using randomized assignment data, as well as nonexperimental studies that assume assignment is independent of u. Discussion of program effects typically centers on mean impacts, and in particular on the mean impact of treatment on the treated, which is the coefficient on D in the regression model just outlined. A more general parameter of interest is $E[\Delta]$, the population average treatment effect.[31]

Assuming Y is a continuous variable, define the joint distribution of (Y_{i0}, Y_{i1}) as $Pr[Y_{i0} \leq y_0, Y_{i1} \leq y_1] \equiv F(y_0, y_1)$. The standard evaluation problem for parameters other than mean impacts is that the pair (Y_{i0}, Y_{i1}) is never observed for any i: it is impossible to observe a person simultaneously in and not in a program. Therefore, the joint distribution is unidentified without

[30] The model we discuss has been called the "Rubin Causal Model" (e.g., Imbens & Rubin (1997)), and the "Roy model" or the "switching regression model" (e.g., Heckman et al. (1997)).

[31] With non-experimental data, no relationship need exist between the effect of treatment on the treated and more general mean-impact parameters. With random assignment of treatment status, the average effect of treatment on the treated and the population average treatment effect are necessarily equal. The one important caveat is that the population for which this statement is valid is limited to the population represented in the experiment. For nonexperimental data, one must assume a functional relationship between the effect of treatment on the treated in order to identify the population average treatment effect. We will not discuss this issue further in this paper, as it has been well-discussed in the literature on program evaluation.

untestable assumptions.

4.2 Treatment effects on distributions

Despite the fact that the joint distribution F is unidentified, the two marginal distributions $Pr(Y_{i1} \leq y_1|D = 1) \equiv F_1(y_1|D = 1)$ and $Pr(Y_{i0} \leq y_0|D = 0) \equiv F_0(y_0|D = 0)$ are always identified, even with nonexperimental data. This is so because any (representative) data source will allow us to trace out the marginal distributions of Y among participants (yielding $F_1(y_1|D = 1)$) and non-participants (yielding $F_0(y_0|D = 0)$). Nonexperimental research necessarily focuses on assumptions and data sufficient to estimate $F_1(y_1|D = 0)$ or $F_0(y_0|D = 1)$, and perhaps $F(y_0, y_1)$.

With random assignment, the marginals $F_0(y_0|D = 0)$ and $F_0(y_0|D = 1)$ are necessarily (almost surely) the same. That is, random assignment ensures the equivalence of the distribution of counterfactual outcomes when not participating for program participants and nonparticipants. Similarly, the marginals $F_1(y_1|D = 1)$ and $F_1(y_1|D = 0)$ are (almost surely) identical given random assignment. That is, random assignment ensures that if they were treated, nonparticipants would have the same marginal distribution as participants do. As an immediate consequence of these equivalences, we have $F_k(y_k|D = j) = F_k(y_k)$ for $j, k \in \{0, 1\}$, with the understanding that the relevant population is limited to women who are statistically "like" the ones in the study (see footnote 31 on the page before). This equivalence means that one can use randomized data to consistently estimate the *treatment effect on the (marginal) distribution* (henceforth TED). This treatment effect is defined as $\Delta F(y) \equiv F_1(y) - F_0(y) = Pr(Y_1 \leq y) - Pr(Y_0 \leq y)$, i.e., the difference in the probability that a randomly chosen person's outcome will not exceed the level y when treated, compared to the probability when not treated.

The QTE is closely related to the TED, as can be seen by inverting $F_k(y) = q$: $y_k(q) = F_k^{-1}(q)$, $k \in \{0, 1\}$.[32] Thus, $y_1(q)$ is the q^{th} quantile of the (marginal) distribution given treatment, and $y_0(q)$ is the same quantile of the (marginal) distribution given no treatment. The QTE parameter is $\Delta y(q) \equiv y_1(q) - y_0(q)$, which is the change in the expected value of the outcome at the q^{th} quantile when we take a randomly-chosen, previously-untreated person and give her the treatment.

The assumption of constant treatment effects implies that the QTE is constant, and that the

[32] In general, $y_k(q)$ is defined as $\inf_y\{y : F_k(y) \geq q\}$. The notation in the text is appropriate if Y has a continuous density with no mass points.

distribution shifts up by uniformly by the mean treatment effect. It is important to emphasize that random, or ignorable, assignment is sufficient to guarantee the consistency of the simple-difference QTE estimator used in section 3. No other assumptions are required. In fact, this is the natural analog to estimating mean impacts in experimental studies by simple differencing means for the treatment and control groups.[33]

Recent studies focusing on estimating TEDs or QTEs include Abadie, Angrist & Imbens (2002), Imbens & Rubin (1997), and Heckman et al. (1997).

4.3 Distributions of treatment effects

Because the joint distribution of (Y_{i0}, Y_{i1}) is never nonparametrically identified, a much more difficult problem than estimating the QTE is to estimate the treatment effect for any given person (or set of persons). For the same reason, estimating the *distribution of treatment effects* (DTE) (defined as $Pr(\Delta_i \leq \delta) \equiv G(\delta)$), or any of its features, requires assumptions other than random assignment. Heckman et al. (1997) discuss a number of measures of interest that depend on the distribution G (or functionals of it). For example, given random assignment, the fraction of "losers" from a program may be written $Pr(\Delta_i < 0) = G(0)$, while the q^{th} quantile of the DTE is given by $\Delta(q) \equiv G^{-1}(q)$. Such parameters are functionals of the joint distribution of potential outcomes and are generally unidentified without assumptions linking the joint and marginal distributions. Much of the discussion in Heckman et al. (1997) concerns the construction of either bounds or other useful but not exhaustive information concerning the distribution of treatment effects. In a future version, we will discuss and make use of some of their methods.

As a brief example of why QTEs and DTEs can differ, consider the following simple example. Suppose there are four people and that their (Y_0, Y_1) pairs are given by $\{(10, 15), (20, 25), (30, 35), (40, 9)\}$. In this case, the list of QTEs is $(-1, -5, -5, -5)$, yet the DTE (as ranked) is $\Delta = (-31, +5, +5, +5)$. Focusing here only on the QTE, one would conclude that the distribution of Y falls everywhere, whereas focusing on the DTE would show that three-fourths of the population had gained a moderate amount, with the other one-fourth absorbing a large loss. The key fact is that *both* of these claims are correct.

[33] As discussed in footnote 24 on page 15, the experimental assignment for Jobs First appears to be non-ignorable, a difficulty we address by using Firpo's (2003) semiparametric QTE estimator. Results from these estimates were qualitatively similar.

This example shows that when treatment leads to rank changes in the marginal distributions, the choice of focusing on QTEs or DTEs can lead to large differences in interpretation. When ranks are preserved, this indeterminacy is resolved. Rank preservation, also known as perfect positive dependence, implies a direct link between the QTE and the DTE, namely that the DTE is given by the order statistics from the QTE. While this assumption nests the assumption of constant treatment effects, it can be quite strong in practice.

To see this, consider Figure 7, which uses the simulation results from section 2 to plot the difference in the *cdf* by earnings quantile when facing AFDC: for a person who starts at quantile q and moves up one decile because of the expanded Jobs First disregard, the point on this graph would be $(q, q+0.1)$. All points would lie on the horizontal line at 0 if there were no rank reversals. Simulated women whose earnings rise under Jobs First have points above this line, and vice-versa. Under the AFDC rules, 36.5% of simulated women have 0 earnings, which explains the absence of any points to the left of the point (0,0.365). The heavy, nearly-straight lines clustered around 0 represent the large majority of women, whose earnings are constant or little-changed with treatment, and whose ranks change very little (rank changes occur for these women because others move either above or below them). The horizontal dotted lines in the figure are drawn to provide a region of plus or minus one decile. Among simulated women who participate in Jobs First but not AFDC, and who have high earnings under Jobs First, a notable fraction lie outside this region. The same is true for women who would participate in AFDC and have zero earnings under AFDC rules. However, more than 94% of simulated differences in earnings quantiles fall within one decile of the original quantile. Thus while full rank preservation is an unattractive assumption based on the simulation results, an assumption that most rank moves are relatively small may not be. From that perspective, thinking of the quantile treatment effects as approximately the treatment effects for people at given quantiles is probably not too unreasonable, provided one realize that a small number of women move quite far in the earnings distribution.

4.4 Discussion: Do names matter?

In light of the discussion above, the appropriate focus of attention in evaluating programs with heterogeneous treatment effects is an important issue. One view is that estimates of treatment effects on distributions are sufficient for evaluation purposes. This is because policy-driven changes affect classical social welfare functions only insofar as they change the marginal distributions of the level of utility. In other words, "names don't matter"; two policies that increase the social welfare function by 100 are viewed equally, even if one reduces utility for all but one person (whose utility rises significantly) and the other raises everyone's utility equally. This approach is discussed in Atkinson (1970) and more recently advocated by Abadie et al. (2002).

An alternative view is that distributions of treatment effects (or some functionals of them) are themselves necessary inputs to the evaluation of programs. For example, Heckman et al. (1997, p. 488) write that "Appeal to a mythical social welfare function begs fundamental questions of political economy. The distribution of the benefits (and costs) from a programme determines the support for a programme if voters are self-interested or if they are altruistic." In other words, "names matter".[34]

In our view, the distribution of treatment effects is an interesting topic of study. However, we believe that the social welfare function approach—namely focusing on quantile treatment effects—hews more closely to the policy questions of interest with respect to welfare reform. Moreover, the political economy questions raised by Heckman et al. (1997) seem relatively unimportant for the welfare population, who are likely to have little political clout, either individually or as a group.

4.5 Assessing the impact of Jobs First using alternative Social Welfare Functions

Future versions of the paper will include evaluations of the reform programs using the marginal distributions, under alternative assumptions on the social welfare function. We plan to use the CRRA family of social welfare functions, suitably normalized, to compare conclusions using these social welfare functions to those using classical cost-benefit mean-impacts analysis.

5 Results for FTP and MFIP Reforms

To be added.

[34] A third approach is that taken by Dehejia (Forthcoming), who treats program evaluation as a decision-theoretic problem. He uses data from the Alameda County portion of California's job search- and training-based GAIN experiment, conducted in the late 1980s, to assess the role of heterogeneity in guiding optimal program assignment in the post-evaluation period. Angrist & Dehejia (2001) take a related approach, allowing for individual and social welfare considerations in the presence of risk and inequality aversion.

6 Conclusion

This paper yields several important results. First, basic labor supply theory predicts that welfare reform should generate systematic heterogeneity in treatment effects. Simulation evidence using parameter estimates from Moffitt's (1983) classic paper shows that this heterogeneity can be significant. This heterogeneity would be difficult to fully characterize using mean impacts. While estimating impacts within subgroups can help, it would not uncover some important results (e.g., the program-induced reduction in earnings at the top of the earnings distribution). From a methodological point of view, a focus on mean impacts thus limits the use and testing of economic theory in the context of program evaluation. A second conclusion is that there is overwhelming evidence of systematic treatment effect heterogeneity in the impact of welfare reform, as measured using Connecticut's Jobs First. Much of this heterogeneity provides strong support to theoretical predictions, and little if any contradicts them.

A final set of conclusions concerns the substantive impact of Jobs First on earnings, transfers, and total (observable) income. The empirical results suggest that expanded earnings disregards can significantly increase earnings of welfare participants. However, high-earning women do appear to reduce earnings in response to the associated expansion of the on-welfare budget set. To our knowledge, ours is the first work using experimental data to demonstrate this result. The results also suggest the possibility that a significant share of the increased transfer payments went to relatively well-off women within the full sample, and may have been essentially inframarginal.

Another important finding concerning earnings is that at the bottom deciles, they do not change—which is to say that they stay at zero. This finding holds whether we consider the full Jobs First sample or the most-disadvantaged subgroup (it is approximately true for the least-disadvantaged as well), and it holds both before and after time limits could bind. Moreover, income at the bottom of the distribution is either unchanged (for the first two years after assignment) or falls (for the second two years). At the same time, transfer income falls in the post-time limit period by an especially large amount at the higher deciles of the transfer income distribution.

Taken as a whole, these findings are cause for concern regarding outcomes for significant numbers of families under Jobs First—and hence similar TANF programs. One of PRWORA's four objectives was to "End the dependence of needy parents on government benefits by promoting job preparation, work, and marriage" (Committee on Ways and Means (2000)). Our results suggest that if a program like Jobs First is to end dependence on public assistance, it is unlikely to do so through "job preparation and work."

Appendix A Details of the Jobs First simulation

Appendix A.1 Preferences used and optimal labor supply

We follow Moffitt (1983), who uses a linear desired-hours function, $h^*(w, N) = \alpha + \beta w + \delta N$. Demographic covariates X and a stochastic term are incorporated by the modeling assumption $\alpha \equiv X\nu + \epsilon$, with $\epsilon \sim N(0, \sigma_\epsilon)$. When $w = 0$, desired hours will still be positive as long as $\alpha + \delta N > 0$. This holds trivially for a woman with 0 nonlabor income and positive α, and more generally if the income effect $|\delta N|$ is sufficiently small given α. Intuitively, this means that preferences are such that labor supply can (locally) fail to be a bad. As a consequence, some women will work positive hours while participating on AFDC.

To see this more formally, consider the direct utility function implied by the linear desired-hours function: $U(H, C) = -ln(\beta - \delta H) - (H - \alpha - \delta C)\delta/(\beta - \delta H)$, where $C = wH + N$ is goods consumption. Since this function is strictly increasing in Y, the slope of indifference curves in (leisure,consumption) space is given by the sign of $MU_H \equiv \partial U/\partial H$ (since $\partial U/\partial H \equiv -\partial U/\partial C$, and leisure is a constant minus H). It can be shown that $\text{sign}(MU_H) = \text{sign}(\alpha + \delta C - H)$. This immediately implies that a given indifference curve can both increase and decrease, and since indifference curves are continuous, any indifference curve that does will be locally horizontal at positive hours. At such choices of hours, the marginal disutility of work, and thus the marginal utility of leisure, is 0. With a 100% tax rate, the price of leisure is 0, so that positive hours with a wage of 0 will be optimal in this case. Our simulation results do yield some women with positive labor supply under AFDC. If substitution effects dominate income effects (which is true with a linear desired-hours function when $\beta > 0 > \delta$, as with Moffitt's estimates), then for such women the increased effective wage due to the expanded disregard will increase hours and earnings on welfare, as with those at the AFDC corner with sufficiently flat indifference curves (steeper indifference curves will yield no change). We note as well that these preferences allow the possibility that indifference curves will have an inverted U-shape, since the sign of indifference curves through points with consumption C may be shown to equal $\text{sign}(\delta\alpha - \beta + \delta^2 C)$, which will be negative if α is sufficiently small (i.e., the woman draws a large negative value of ϵ). Therefore, it is important to check the second-order condition to ensure that a utility maximum, rather than a minimum, is achieved by satisfying the first-order condition. We do this by making global comparisons of the corner bundles with interior values of hours implied by the linear labor supply function. Among welfare participants, a nontrivial number have concave indifference curves and wind up at the corner.

We proceed by computing optimal hours on each budget segment, imposing corner solutions where necessary (for example, we set optimal on Jobs First to FPL/w for a woman having $\alpha + \beta w + \delta(G + N) > FPL/w$). We then use the direct utility function to compute utility at each of three possible choices of hours: the optimal hours given no welfare participation, the optimal hours (if positive) given welfare participation, and the zero-hours consumption bundle on welfare, which is $(0, G)$. The optimal choice is the one with maximum utility. In making this calculation, we include the stigma term ϕ in utility on welfare. This term is modeled as $0.65 + u$, where 0.65 is Moffitt's estimate and $u \sim N(0, 1)$, as assumed by Moffitt. Setting the standard deviation to 1 is a normalization, not a substantive assumption, since cardinal utility values are not observable (or defined). Note that u and ϵ are independent of each other, so that relative to their expectations, particular combinations of stigma and hours can both be large, both be small, or be large and small (and vice-versa).

Appendix A.2 Parameter values used

The covariate means we use come from Moffitt's Table 1, on p. 1028 of his paper. The estimated hours equation coefficients come from his "Set 1" estimates in Table 2, p. 1031. They are (means and hours equation coefficient estimates in parentheses): age (35.6, 0.12); family size (3.3, 3.30) number of children (2.1, -4.90); years of schooling (11.5, 3.30) dummy for being nonwhite (0.39, -5.01); local unemployment rate (9.0, -2.70); constant (1, -5.01). To get each observation's value of the intercept α in the linear labor supply equation, we take the inner product of these means and their coefficient estimates and then add a random disturbance drawn from a $N(0, 28.50)$ distribution (where 28.50 is Moffitt's estimated standard deviation for this disturbance). The other coefficient estimates for the hours equation are 3.10 for the wage and -0.20 for nonlabor income.

Moffitt argues that the effective marginal tax rate on AFDC is substantially lower than 100% due to deductions (e.g., for child care and work expenses). Taking account of such deductions adds another segment of the budget set on AFDC, so we ignore them for simplicity; moreover, if child care and transportation expenditures have no *per se* value as consumption goods, then deductions for them are appropriately ignored.

Stata code that generates the simulation is available on request.

References

Abadie, A., Angrist, J. & Imbens, G. (2002), 'Instrumental variables estimates of the effect of subsidized training on the quantiles of trainee earnings', *Econometrica* .

Angrist, J. & Dehejia, R. (2001), When is ATE enough? Risk aversion and inequality aversion in evaluating training programs. Typescript.

Atkinson, A. B. (1970), 'On the measurement of inequality', *Journal of Economic Theory* **2**, 244–263.

Bane, M. J. & Ellwood, D. T. (1994), *Welfare Realities: From rhetoric to reform*, Harvard University Press, Cambridge and London.

Bennett, N., Lu, H.-H. & Song, Y. (2002), Welfare reform and changes in the economic well-being of children, Working Paper 9399, NBER.

Bitler, M. P., Gelbach, J. B. & Hoynes, H. W. (2003, (Papers and Proceedings)), 'Some evidence on race, welfare reform and household income', *American Economic Review* **93**(2).

Blank, R. M. (2002), Evaluating welfare reform in the United States, Working Paper 8983, NBER.

Bloom, D., Kemple, J. J., Morris, P., Scrivener, S., Verma, N. & Hendra, R. (2000), *The Family Transition Program: Final Report on Florida's Initial Time-Limited Welfare Program*, Manpower Demonstration Research Corporation, New York, NY.

Bloom, D. & Michalopoulos, C. (2001), How did welfare and work policies affect employment and income: A synthesis of research, Working paper, Manpower Demonstration Research Corporation.

Bloom, D., Scrivener, S., Michalopoulos, C., Morris, P., Hendra, R., Adams-Ciardullo, D. & Walter, J. (2002), *Jobs First: Final Report on Connecticut's Welfare Reform Initiative*, Manpower Demonstration Research Corporation, New York, NY.

Cabrera, N. & Evans, V. J. (2000), 'Welfare reform and its consequences: What questions are left unanswered?', *Joint Center for Poverty Research Poverty Research News* **4**(6).

Committee on Ways and Means, U. (2000), *Background Materials and Data on Programs Within the Jurisdiction of the Committe on Ways and Means*, U.S. Government Printing Office, Washington.

Dehejia, R. H. (Forthcoming), 'Program evaluation as a decision problem', *Journal of Econometrics* .

Firpo, S. (2003), Efficient semiparametric estimation of quantile treatment effects. Typescript.

Grogger, J. (Forthcoming), 'The effect of time limits, the EITC, and other policy changes on welfare use, work, and income among female-headed families', *Review of Economics and Statistics* .

Grogger, J., Haider, S. & Klerman, J. (2003 (Papers and Proceedings)), 'Why did the welfare rolls fall during the 1990s? the importance of entry', *American Economic Review* .

Grogger, J., Karoly, L. A. & Klerman, J. A. (2002), Consequences of welfare reform: A research synthesis, Working Paper DRU-2676-DHHS, RAND.

Grogger, J. & Michalopoulos, C. (Forthcoming), 'Welfare dynamics under time limits', *Journal of Political Economy* .

Heckman, J. J., Smith, J. & Clements, N. (1997), 'Making the most out of programme evaluations and social experiments: Accounting for heterogeneity in programme impacts', *Review of Economic Studies* **64**, 487–535.

Imbens, G. W. & Rubin, D. B. (1997), 'Estimating outcome distributions for compliers in instrumental variables models', *Review of Economic Studies* **64**, 555–574.

Michalopoulos, C. & Schwartz, C. (2000), What works best for whom: Impacts of 20 welfare-to-work programs by subgroup, Working paper, Manpower Demonstration Research Corporation, U.S. Department of Health and Human Services, and U.S. Department of Education.

Miller, C., Knox, V., Gennetian, L. A., Dodoo, M., Hunter, J. A. & Redcross, C. (2000), *Reforming Welfare and Rewarding Work: Final Report on the Minnesota Family Investment Program Volume 1: Effects on Adults*, Manpower Demonstration Research Corporation, New York, NY.

Moffitt, R. (1983), 'An economic model of welfare stigma', *American Economic Review* **73**(5), 1023–35.

Moffitt, R. (1999), The effect of pre-PRWORA waivers on welfare caseloads and female earnings, income and labor force behavior, *in* S. Danziger, ed., 'Economic Conditions and Welfare Reform', W. E. Upjohn Institute for Employment Research, Kalamazoo, MI.

Moffitt, R. (2002), Welfare programs and labor supply, Working Paper 9168, NBER.

Schoeni, R. & Blank, R. (2003 (Papers and Proceedings)), 'Changes in the distribution of child well-being over the 1990s', *American Economic Review* **93**(2).

Schoeni, R. F. & Blank, R. M. (2000), What has welfare reform accomplished? Impacts on welfare participation, employment, income, poverty, and family structure, Working Paper 7627, NBER.

Figure 1: Stylized Connecticut Budget Constraint Under AFDC and Jobs First

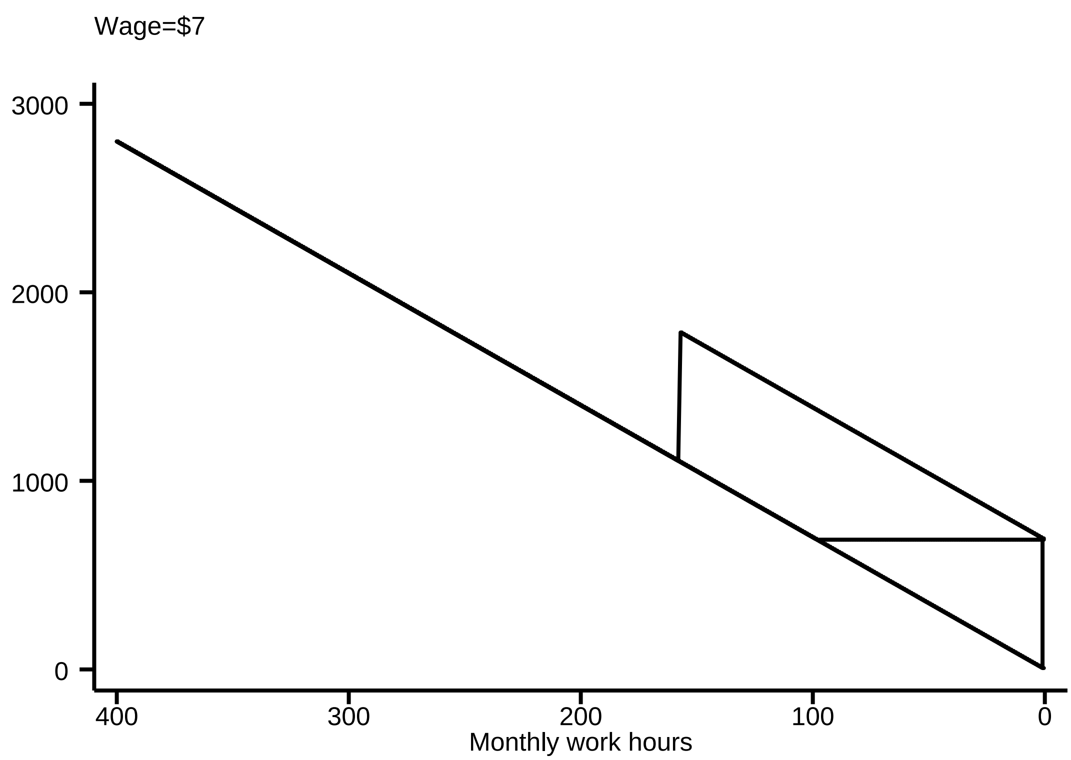
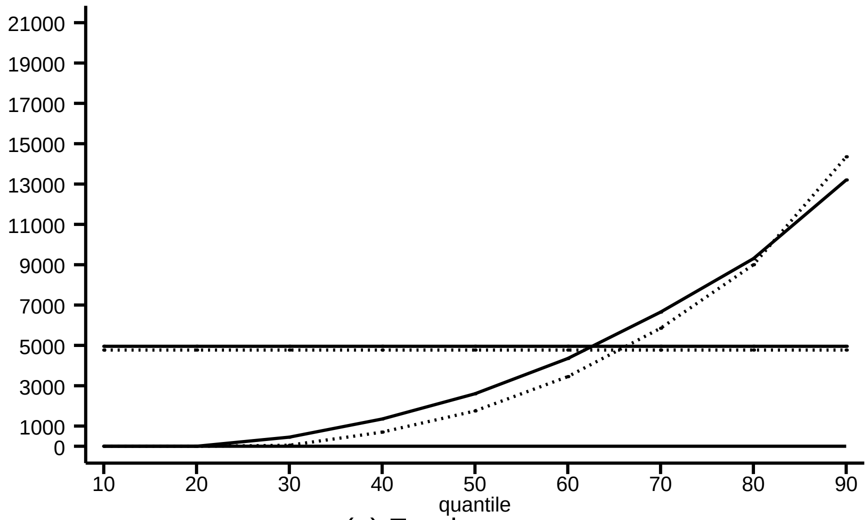
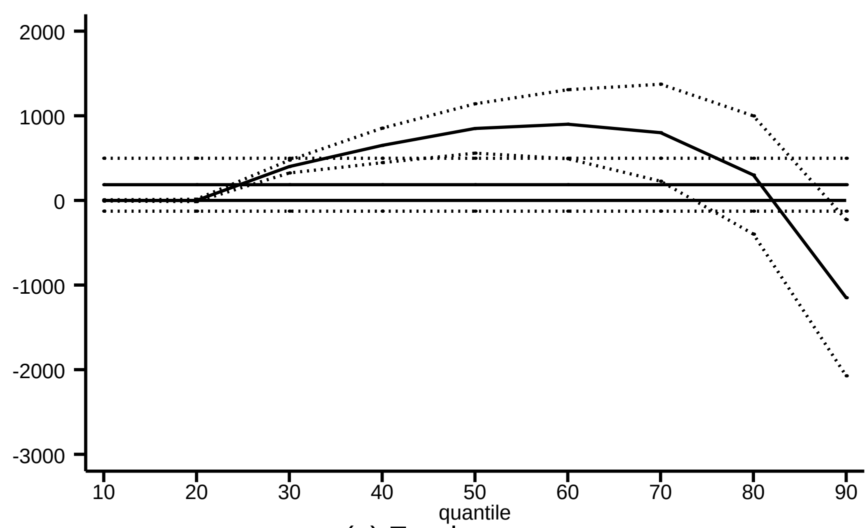


Figure 2: Jobs First Income Distributions and Treatment Effects on Quantiles, Years 1–2



(a) Earnings

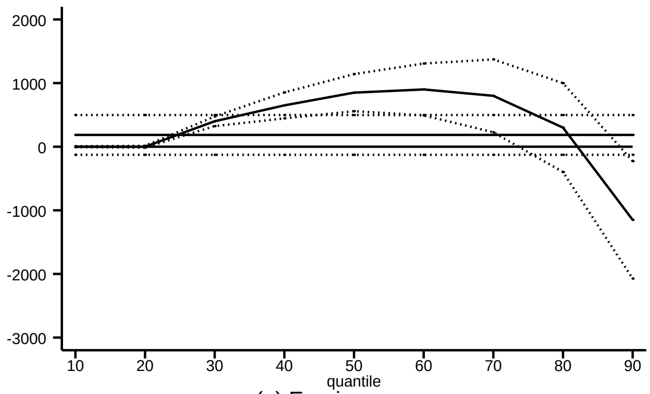


(a) Earnings

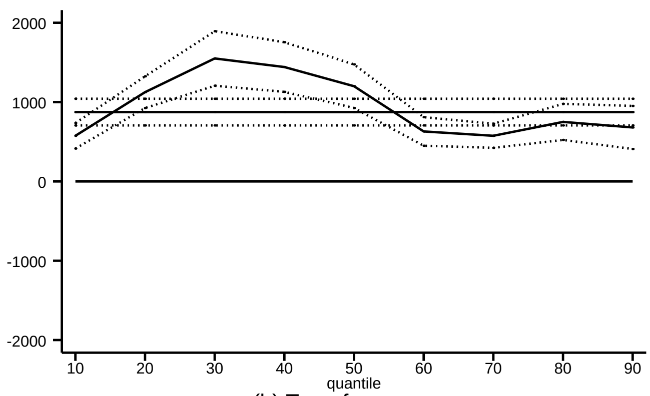
Dotted lines⇒Control Group
Solid lines⇒Treatment Group

Figure 3: Jobs First Treatment Effects on Quantiles (all observations, no controls)

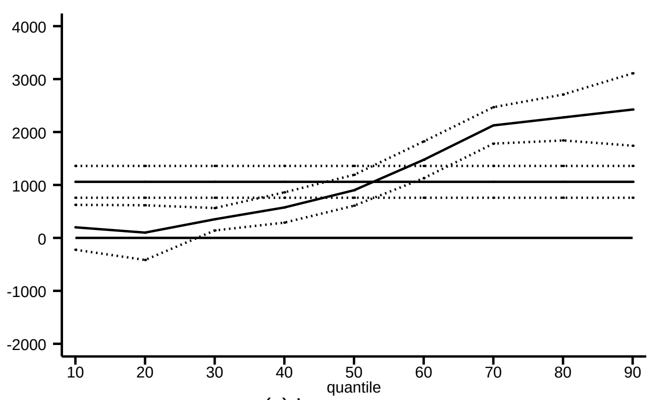
Quarters 1–8



(a) Earnings

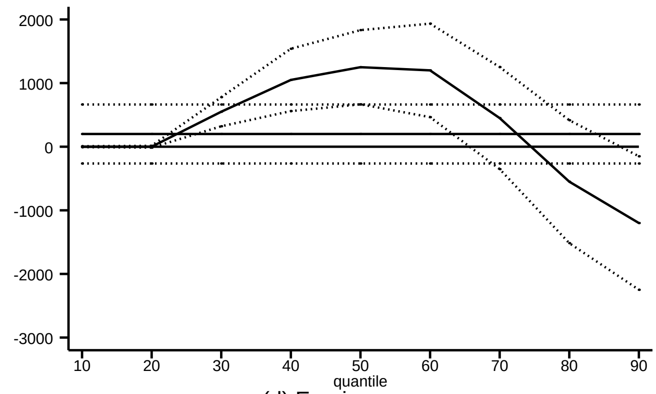


(b) Transfers

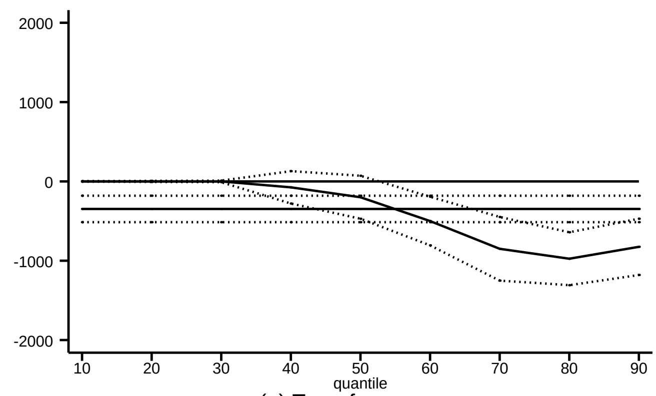


(c) Income

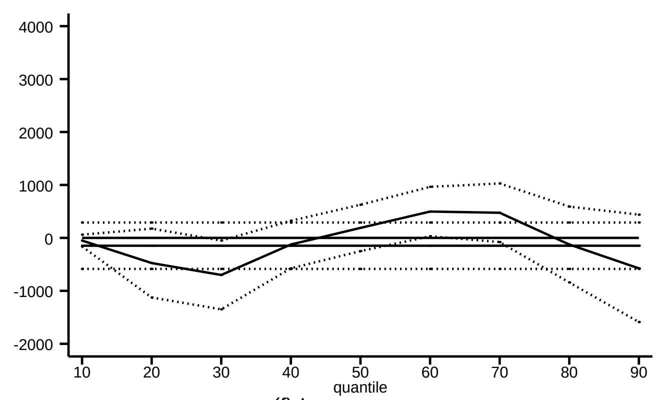
Quarters 9–16



(d) Earnings

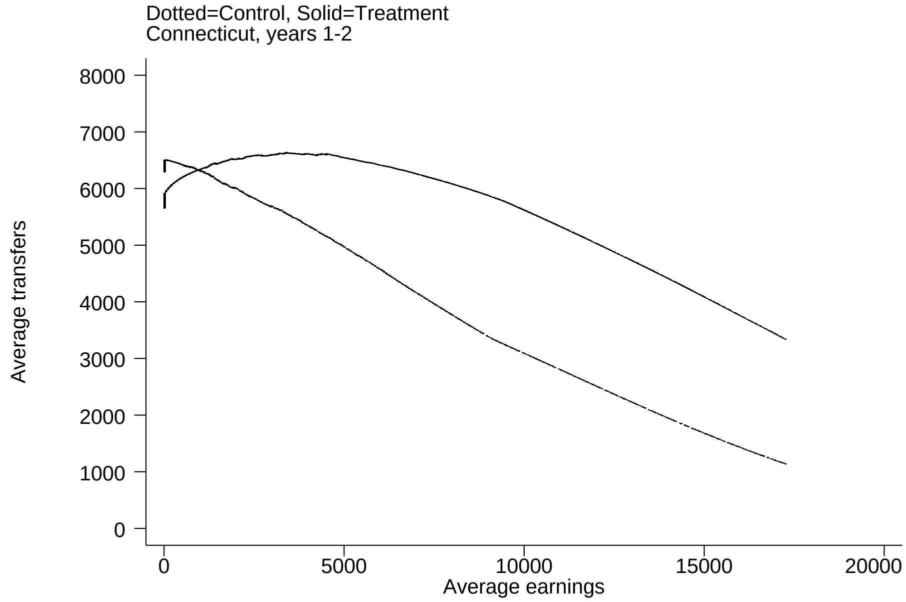


(e) Transfers



(f) Income

Figure 4: Lowess Regressions of Transfers on Earnings, Connecticut



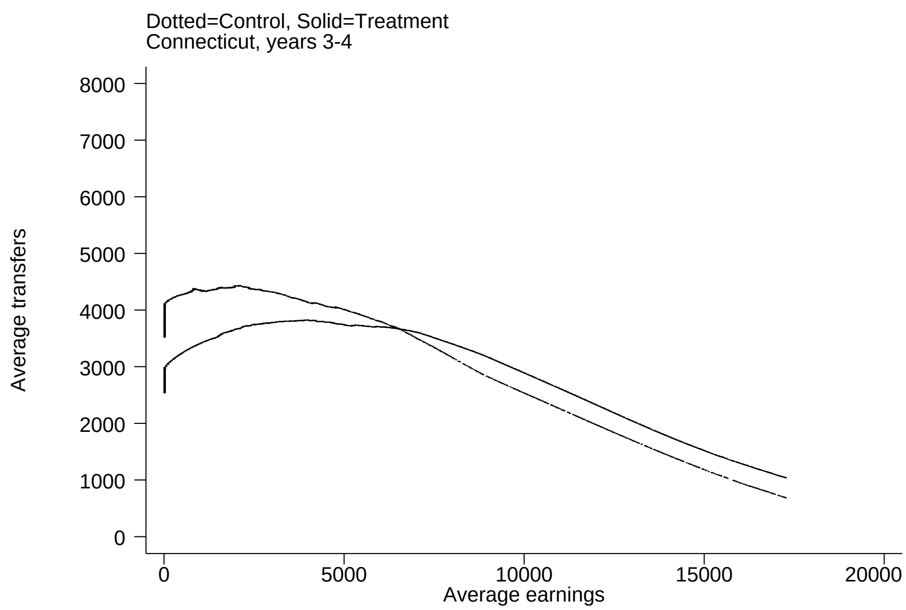


Figure 5: Jobs First Treatment Effects on Quantiles (LD observations, no controls)

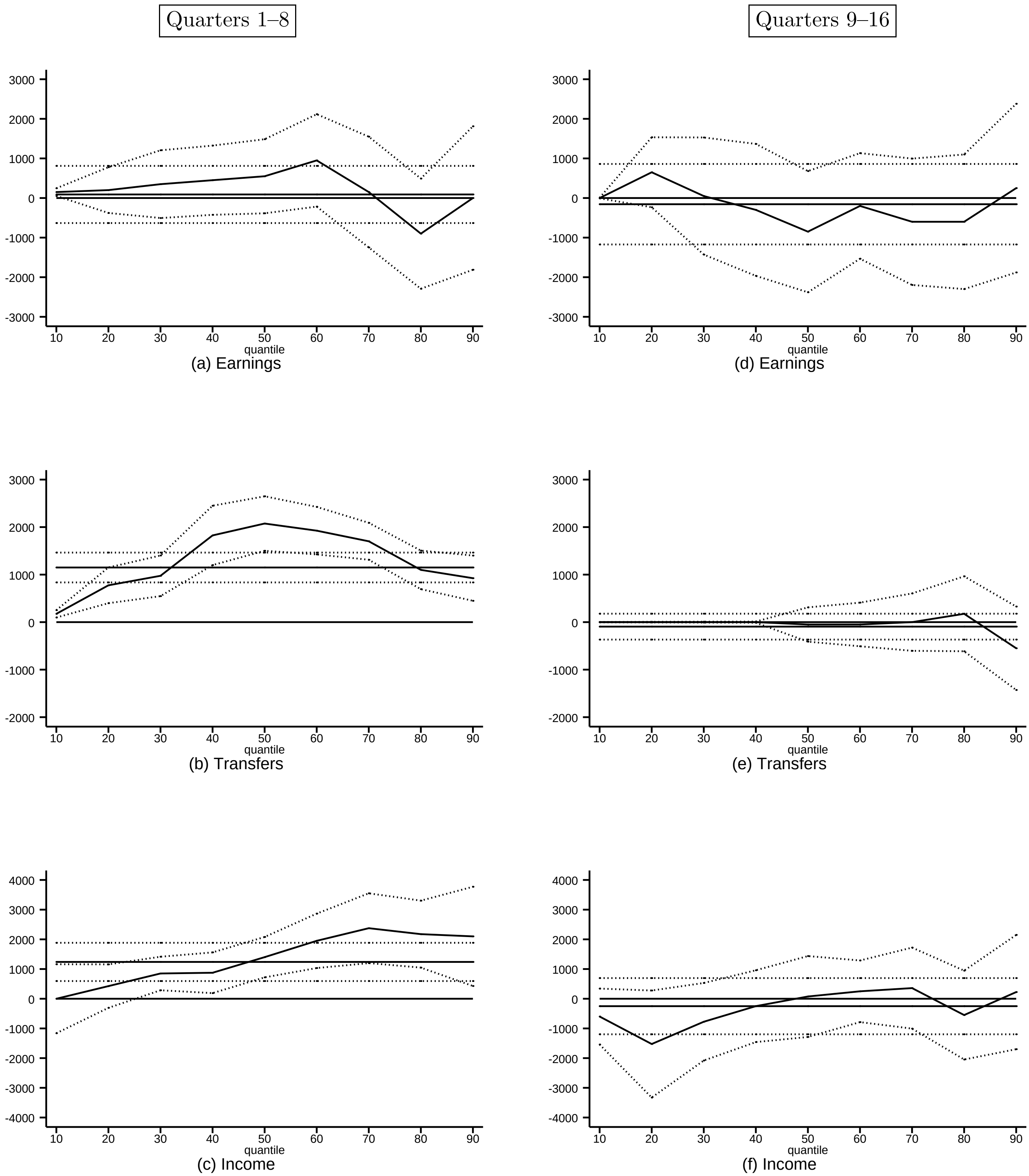


Figure 6: Jobs First Treatment Effects on Quantiles (most disadvantaged observations, no controls)

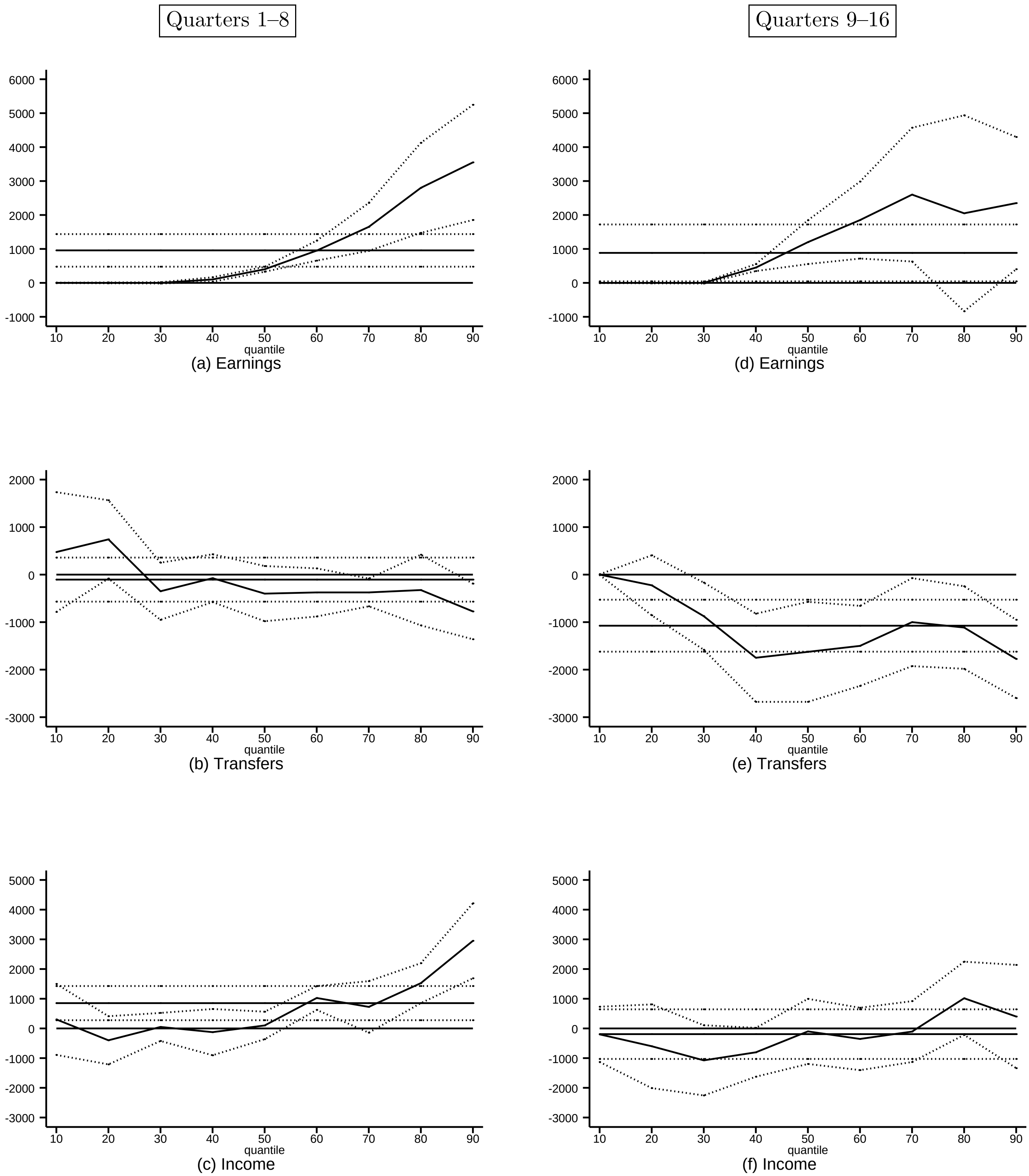


Figure 7: Simulated Difference in cdf Value, by Simulated AFDC Earnings Quantile

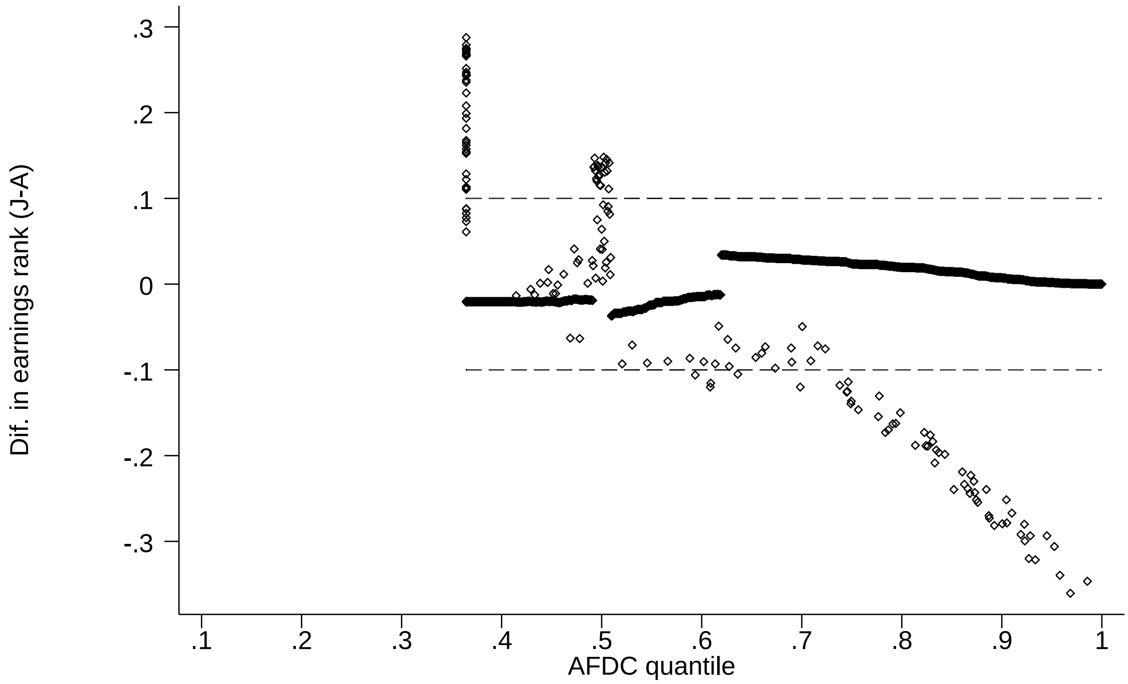


Table 1: Simulated Outcomes Using Moffitt (1983) Estimates

			AFDC Non-participants			
	AFDC Participants		All		JF participants only	
	AFDC rules	JF Impact	AFDC rules	JF Impact	AFDC rules	JF Impact
Welfare participation	1.000	0.000	0.000	0.054	0.000	1.000
Hours	2.05	2.34	29.12	-1.32	47.83	-24.59
Earnings	2.08	3.74	58.94	-2.55	83.34	-47.46
Welfare income	19.41	2.08	0.00	1.15	0.00	21.49
Total income	21.49	5.82	58.94	-1.39	83.34	-25.96
Pop. share (AFDC participation rate)	0.246		0.754		0.041	

AFDC Participants are those who would participate in welfare if facing AFDC rules; *AFDC Non-participants* are those who would not participate if facing AFDC rules. Among *AFDC Non-participants*, the *JF Participants only* are those who would participate in welfare only if assigned to Jobs First. Figures in JF Impact column for last group thus equal figures for all *AFDC Non-participants*, divided by the treatment effect on welfare participation for all *AFDC Non-participants*. Wages expressed as multiple of the minimum wage (which was $2.10 in 1975). Simulations use lognormal wage distribution with median of 1.56 minimum wage units. Maximum weekly welfare-plus-Food Stamps benefit is 21.49 minimum wage units.

Table 2: Policy Changes in Waiver and Control AFDC Programs

	Connecticut Jobs First	Florida FTP	Minnesota MFIP	Pre-existing AFDC
Time limit	*Stringent* 21 months (6-month extension if in compliance and income less than benefits)	*Moderate* 24 months in any 60 month period, 36 in any 72 month period for least job-ready	*None*	*None*
Work Requirements	*Moderate* Mandatory work first, exempt if child < 1	*Moderate* Job search for job ready, education/training for non-job-ready, exempt if child < 6 months	*Strong* Strong employment (mandatory only for long-term recipient), exempt if child < 1	*Weak* Education/training, exempt if child< 2 (CT), < 3 (FL, MN)
Sanctions	*Mild* 1^{st}: 20% cut, 3 months 2^{nd}: 35% cut, 3 months 3^{rd}: grant cancelled, 3 months	*Moderate* Adult removed from grant until compliant	*Weak* 10% cut	*Minimal (rarely enforced)* 1^{st}: adult removed from grant until compliant 2^{nd}: adult removed ≥ 3 months 3^{rd}: adult removed ≥ 6 months
Earnings Disregard	*Generous* All earned income disregarded up to poverty line	*Moderate* \$200 + 50% of remaining earnings	*Moderate* 38% of earnings disregarded up to 140% of poverty line, maximum grant increased by 20% if working	*Minimal* Months 1–3: \$120+1/3 Months 4–12: \$120 Months > 12: \$90
Other policies	•Asset limit \$3000 •Partial family cap (50%) •2 years transitional Medicaid •Child care assistance •Child support pass-through	•Asset limit \$5000	•Asset limit \$2000 •Eliminate 100-hour & work requirement for 2-parent families •Other relative's income counted in benefit formula •Food stamp cash out •Streamlined administrative procedures	•Asset limit \$1000 •100-hour rule and work history requirement for 2-parent families •1-year transitional Medicaid

Sources: Bloom et al. (2002), Bloom et al. (2000), and Miller et al. (2000).

Table 3: Overview of Evaluation and Sample Sizes

	Connecticut Jobs First	Florida FTP	Minnesota MFIP
Timing	Assignment: 1/96–2/97 Follow-up: 4 years (16 quarters)	Assignment: 5/94–2/95 Follow-up: 4 years (16 quarters)	Assignment: 4/94–3/96 Follow-up: 2-3 years (depends on data source)
Geographic Range	Statewide waiver program Evaluation in 2 offices (Manchester & New Haven)	Partial state waiver Evaluation for Escambria county (includes Pensacola)	Partial state waiver Evaluation for 7 counties, including Hennepin (includes Minneapolis)
Sample sizes	4,803 single parent cases Treatment: 2,396 Control: 2,407	2,815 single parent cases Treatment: 1,405 Control: 1,410	9,217 single parent cases Urban/long-term: Treatment: 846 Control: 934 Urban/applicants Treatment: 1,916 Control: 2,133

Sources: Bloom et al. (2002), Bloom et al. (2000), and Miller et al. (2000).

Table 4: Means for Jobs First and the 1994 National Caseload

	Full sample		Most disadvantaged		Least disadvantaged		National
	Treatment	Control	Treatment	Control	Treatment	Control	Caseload
Applicant sample	0.376	0.407	0.033	0.040	0.655	0.673	
Worked previous year	0.502	0.542	0.000	0.000	1.000	1.000	0.387
Any AFDC previous year	0.669	0.638	1.000	1.000	0.416	0.402	1.000
Race/ethnicity							
White	0.362	0.348	0.298	0.324	0.503	0.404	0.450
Black	0.368	0.371	0.365	0.328	0.393	0.411	0.367
Hispanic	0.207	0.216	0.311	0.312	0.080	0.159	0.131
Marital status							
Never married	0.654	0.661	0.721	0.673	0.635	0.666	0.444
Div/wid/sep/living apart	0.332	0.327	0.266	0.327	0.355	0.325	0.343
Education							
HS dropout	0.350	0.334	0.923	0.956	0.000	0.000	0.374
HS diploma/GED	0.583	0.604	0.000	0.000	1.000	1.000	0.377
More than HS diploma	0.063	0.058	0.077	0.044	0.000	0.000	0.249
N	2,396	2,407	299	250	473	560	765

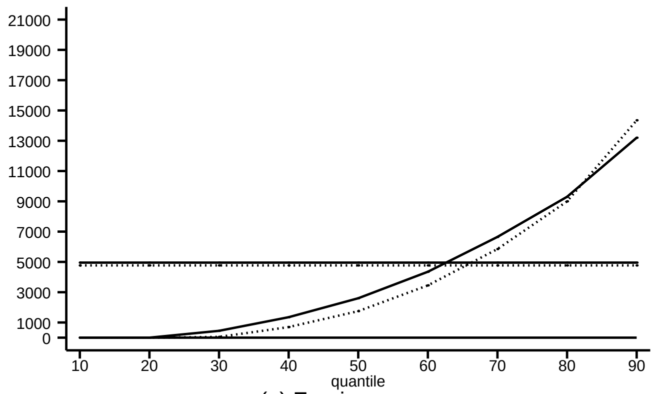
Note: National caseload statistics were constructed using all females aged 16–54 in the 1995 March CPS who had an own child in the household and whose family was reported to have positive AFDC income for calendar year 1994. All national caseload statistics are computed using March supplementary weights. Standard deviations omitted because all variables are binary.

Table 5: Jobs First Mean Impacts

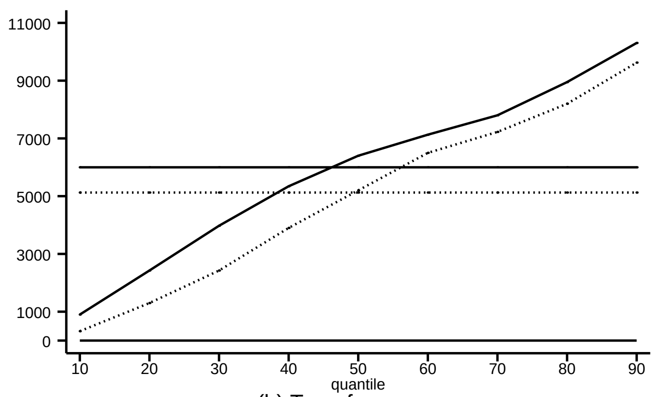
	Years 1–2		Years 3–4	
	Simple Differences	Regression-Adjusted	Simple Differences	Regression-Adjusted
Full sample				
Income	1,059***	1,144***	-147	32
	(183)	(154)	(266)	(238)
Earnings	185	418***	200	491*
	(190)	(156)	(282)	(250)
Transfers	874***	725***	-347***	-458***
	(102)	(90)	(102)	(95)
N	4,803	4,803	4,773	4,773
Most disadvantaged				
Income	852**	915***	-192	-164
	(350)	(311)	(508)	(486)
Earnings	957***	936***	883*	862*
	(292)	(293)	(510)	(512)
Transfers	-105	-22	-1,075***	-1,026***
	(282)	(227)	(333)	(307)
N	549	549	544	544
Least disadvantaged				
Income	1,239***	1,232***	-251	-175
	(392)	(353)	(575)	(554)
Earnings	90	124	-157	-82
	(439)	(387)	(618)	(591)
Transfers	1,150***	1,108***	-95	-93
	(191)	(185)	(166)	(164)
N	1,033	1,033	1,030	1,030

Appendix Figure 1: Jobs First Treatment and Control Group Marginal Distributions (all observations, no controls)

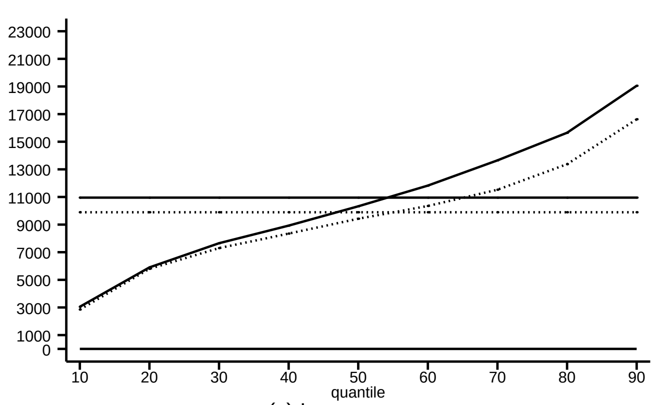
Quarters 1–8



(a) Earnings

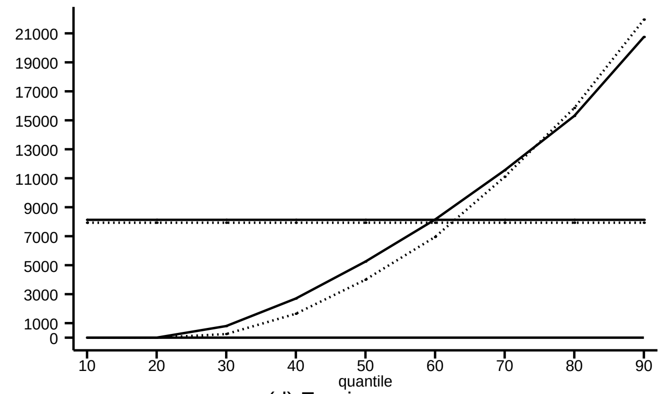


(b) Transfers

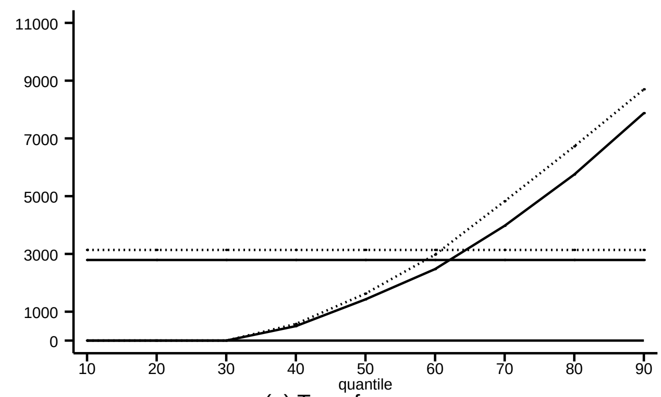


(c) Income

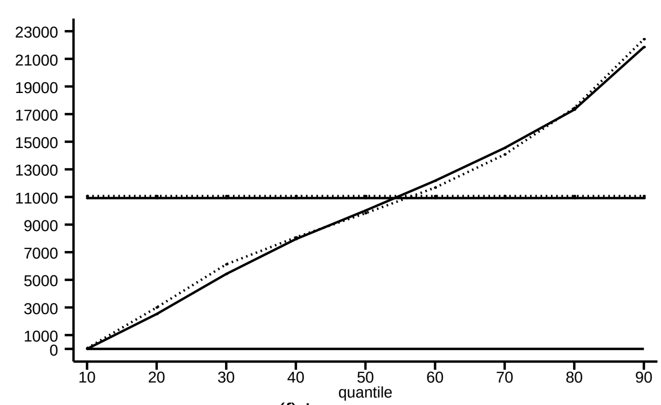
Quarters 9–16



(d) Earnings



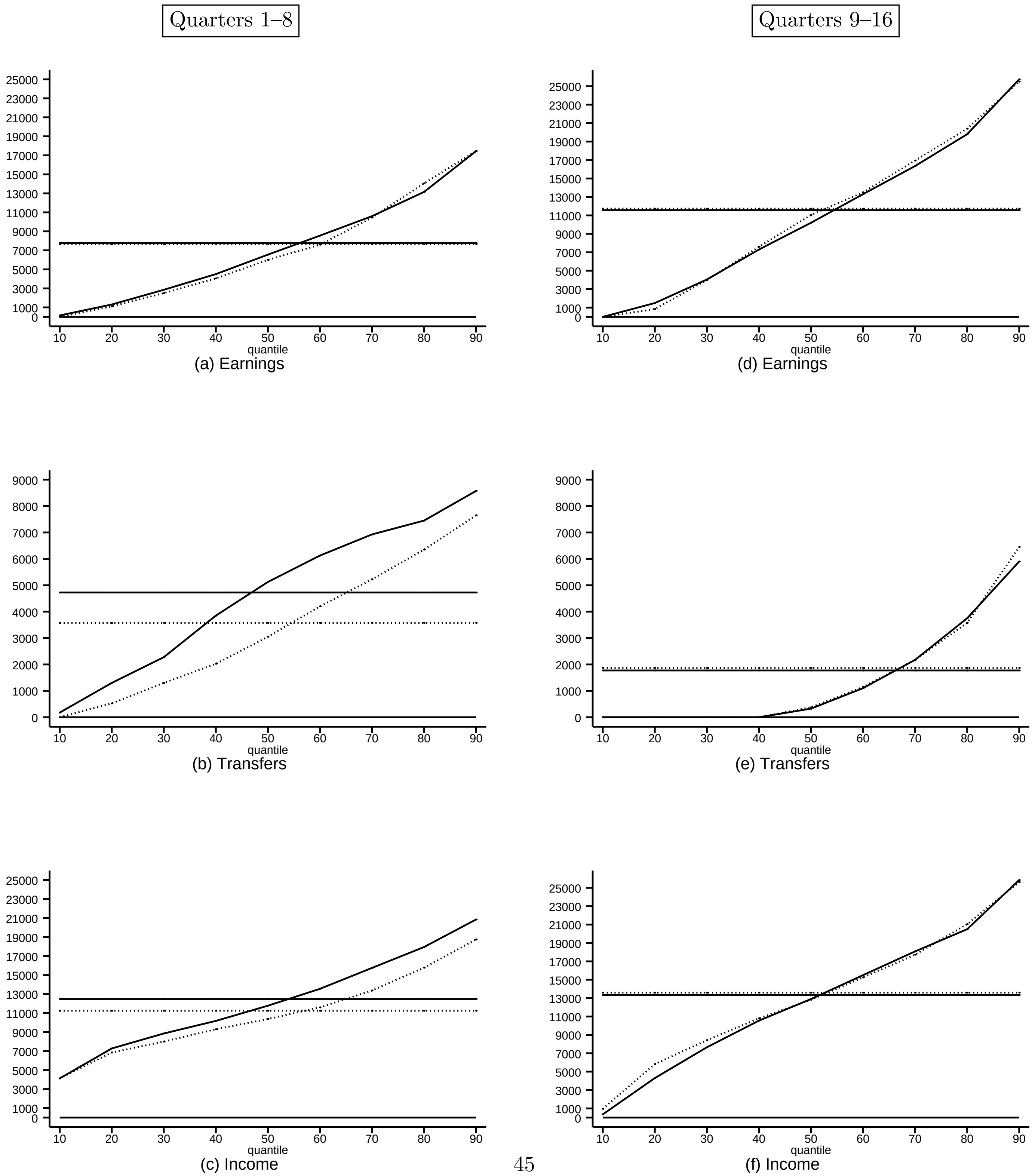
(e) Transfers



(f) Income

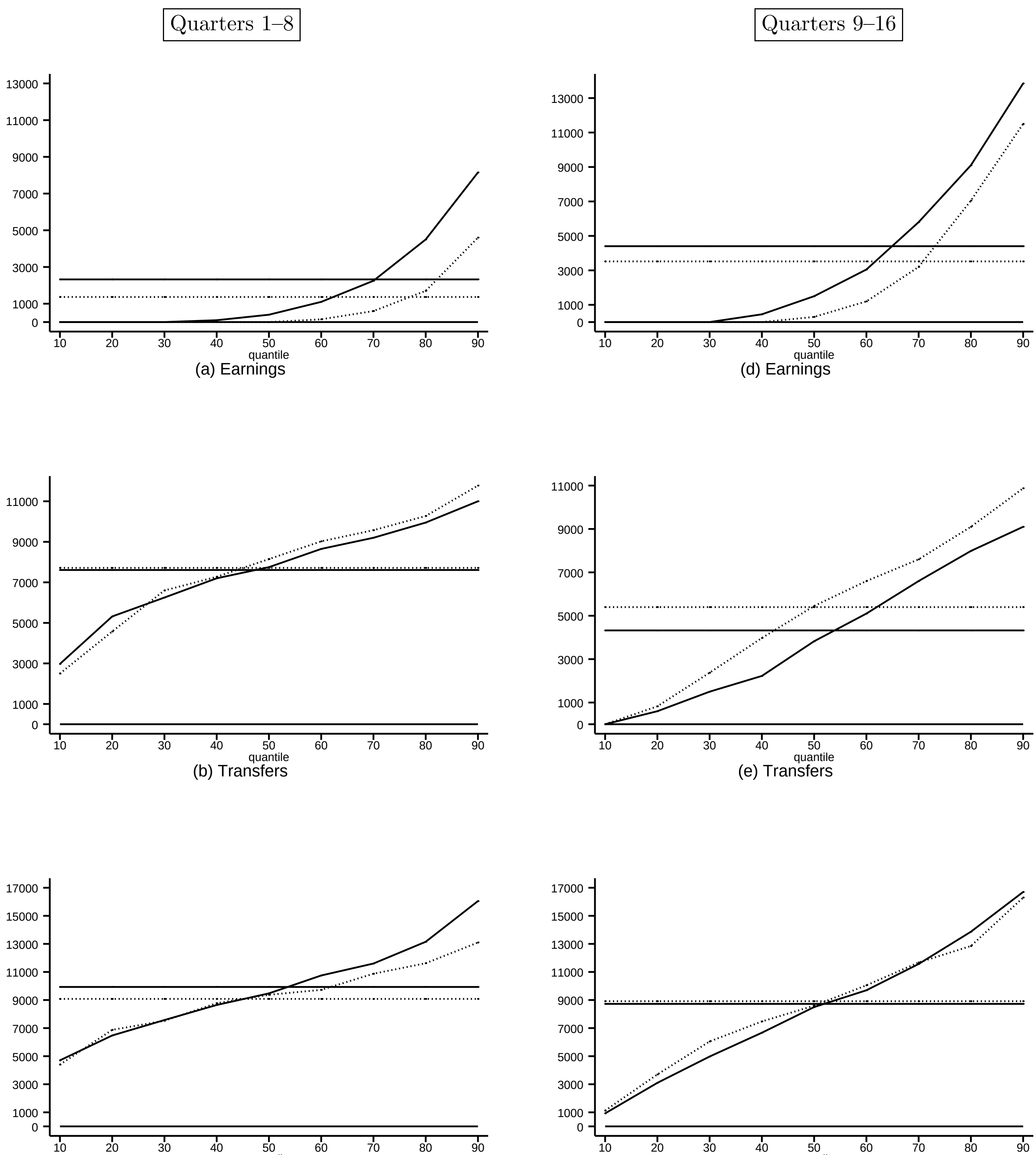
Dotted lines⇒Control Group
Solid lines⇒Treatment Group

Appendix Figure 2: Jobs First Treatment and Control Group Marginal Distributions (LD observations, no controls)



Dotted lines⇒Control Group
Solid lines⇒Treatment Group

Appendix Figure 3: Jobs First Treatment and Control Group Marginal Distributions (most disadvantaged observations, no controls)



Dotted lines⇒Control Group
Solid lines⇒Treatment Group