What Mean Impacts Miss: Distributional Effects of Welfare Reform Experiments

By MARIANNE P. BITLER, JONAH B. GELBACH, AND HILARY W. HOYNES*

Labor supply theory predicts systematic heterogeneity in the impact of recent welfare reforms on earnings, transfers, and income. Yet most welfare reform research focuses on mean impacts. We investigate the importance of heterogeneity using random-assignment data from Connecticut's Jobs First waiver, which features key elements of post-1996 welfare programs. Estimated quantile treatment effects exhibit the substantial heterogeneity predicted by labor supply theory. Thus mean impacts miss a great deal. Looking separately at samples of dropouts and other women does not improve the performance of mean impacts. We conclude that welfare reform's effects are likely both more varied and more extensive than has been recognized. (*JEL* D31, I38, J31)

Nearly a decade 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. For example, TANF programs include lifetime limits on program participation, enhanced work incentives through expanded earnings disregards, stringent work requirements, and financial sanctions for failure to comply with these requirements.

In evaluating the economic effects of welfare reform, it is of first-order importance to assess how reform affects family earnings and income. We start with the simple observation that theory makes heterogeneous predictions concerning the sign and magnitude of the response of labor supply and welfare use to these reforms. Notwithstanding this observation, the vast majority of welfare reform studies rely on estimating mean impacts. Theory predicts that these mean impacts will average together positive and negative labor supply responses, possibly obscuring the extent of welfare reform's effects. Therefore, a critical element in evaluating recent dramatic changes in welfare policy is to measure TANF's impact on earnings and income in a way that allows for heterogeneous treatment effects. That is the focus of this study.

An enormous welfare reform literature has developed in the last several years. We confine our discussion of this literature to a few particularly relevant papers; excellent comprehensive summaries of the literature appear in

* Bitler: Public Policy Institute of California, 500 Washington Street, Suite 800, San Francisco, CA 94111 (e-mail: bitler@ppic.org); Gelbach: Department of Economics, University of Maryland at College Park, College Park, MD 20742, and Florida State University College of Law (e-mail: gelbach@glue.umd.edu); Hoynes: Department of Economics, University of California, Davis, 1152 Social Science and Humanities Building, One Shields Avenue, Davis, CA 95616 (e-mail: hwhoynes@ucdavis.edu). Bitler gratefully acknowledges the financial support of the National Institute of Child Health and Human Development, the National Institute on Aging, and the RAND Corporation. 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 or PPIC. The data used in this paper are derived from data files made available to researchers by the Manpower Demonstration Research Corporation (MDRC). The authors remain solely responsible for how the data have been used or interpreted. We are very grateful to MDRC for providing the public access to the experimental data used here. We would also like to thank two anonymous referees, Dan Bloom, Mary Daly, Jeff Grogger, Richard Hendra, Guido Imbens, Sanders Korenman, Chuck Michalopoulos, Lorien Rice, Jesse Rothstein, Susan Simmat, Jeff Smith, Till von Wachter, Arthur van Soest, and Johanna Walter for helpful conversations, as well as seminar participants from the University of California, Berkeley, University of Chicago Harris School, Cornell University, University of California, Davis, University of Delaware, George Washington University, the IRP Summer Research Workshop, Johns Hopkins University, University of Maryland, the National Bureau of Economic Research, Population Association of America, PPIC, the RAND Corporation, Society of Labor Economists, and Syracuse University.

reviews by Rebecca M. Blank (2002), Robert A. Moffitt (2002), and Jeffrey T. Grogger and Lynn A. Karoly (2005). Nonexperimental studies (e.g., Moffitt, 1999, and Grogger, 2003) have found mixed results concerning the impact of welfare reform on income. Experimental studies examining pre-PRWORA state reforms suggest that generous increases in earnings disregards are important for generating mean income gains, but that these gains disappear after time limits take effect (e.g., Dan Bloom and Charles Michalopoulos, 2001, Grogger and Karoly, 2005). With respect to treatment effect heterogeneity, Blank and Robert F. Schoeni (2003) use Current Population Survey (CPS) data to compare the full distribution of the income-to-needs ratio before and after TANF, finding increases at all but the very lowest percentiles. As they discuss, however, their simple before-and-after methods cannot distinguish impacts of TANF from the influence of strong labor markets. The most common way to address distributional concerns is to estimate 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.[1] Michalopoulos and Christine Schwartz (2001) review 20 randomized experiments and conclude, "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 and Karoly (2005) 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).

In this paper, we address heterogeneous theoretical predictions by estimating quantile treatment effects (QTE) across the distributions of earnings, transfer payments (cash welfare plus food stamps), and total measurable income (the sum of earnings and transfers). This method allows us to test, for example, whether the impact of reform is constant across the distribution, or whether reform leads to larger changes in earnings in some parts of the distribution.[2] We examine impacts across the distribution using public-use data files from the Manpower Demonstration and Research Corporation's (MDRC) experimental evaluation of Connecticut's Jobs First waiver from AFDC rules. Our choice to use experimental data, and the Jobs First program in particular, is not incidental. First, as discussed in Blank (2002) and formalized in Bitler et al. (2003a), identifying the impact of TANF using nonexperimental methods is difficult given that TANF was implemented in all states within a very short period and during the strongest economic expansion in decades. Having access to experimental data is particularly useful because it allows us to investigate treatment effect heterogeneity in a context where the source of identification is clear. Second, the Jobs First program (which we discuss in detail below) has both the most generous earnings disregard in the nation and the strictest time limit. It thus provides ideal terrain for investigating whether theoretically predicted treatment effect heterogeneity actually occurs.

Our empirical findings may be summarized with four important conclusions. First, we find evidence of substantial heterogeneity in response to welfare reform. Second, the heterogeneity is broadly consistent with the predictions of static labor supply theory. We find that Jobs First had no impact on the bottom of the earnings distribution, it increased the middle of the earnings distribution, and—before time limits took effect—it reduced the top of the earnings distribution. Third, contrary

[1] Schoeni and Blank (2000) compare the twentieth and fiftieth percentiles of the CPS family income distribution before and after implementation of TANF. They find negative (but insignificant) impacts of TANF on the twentieth percentile, and positive and significant impacts on the fiftieth percentile for a sample of women with less than a high-school education. Some of the MDRC waiver evaluations (e.g., Bloom et al., 2002, and Bloom et al., 2000) include estimates comparing the fraction of treatment and control group members with income in broad categories. This approach, essentially a tabular form of histogram plots, is similar in spirit to ours.

[2] QTE have been used in previous experimental evaluations. Examples of their use in evaluating the Job Training and Partnership Act include James J. Heckman et al. (1997), Sergio Firpo (2005), and Alberto Abadie et al. (2002); Daniel Friedlander and Philip K. Robins (1997) estimate QTE in evaluating effects of job training in earlier welfare reform experiments. The source of heterogeneous treatment effects in these cases is difficult to identify, however, since they mostly involve changes to training programs or job search assistance. Unlike such black-box reforms, in the present context it is clear at least in part why theoretical predictions are heterogeneous.

TABLE 1—KEY DIFFERENCES IN JOBS FIRST AND AFDC PROGRAMS

	Jobs First	AFDC
Earnings disregard	All earned income disregarded up to poverty line (policy also applied to food stamps)	Months 1–3: \$120 + 1/3 Months 4–12: \$120 Months > 12: \$90
Time limit	21 months (6-month extension if in compliance and nontransfer income less than maximum benefit)	*None*
Work requirements	Mandatory work first, exempt if child < 1	Education/training, exempt if child < 2
Sanctions	1st violation: 20-percent cut for 3 months 2nd violation: 35-percent cut for 3 months 3rd violation: grant cancelled for 3 months	(*Rarely enforced*) 1st: adult removed from grant until compliant 2nd: adult removed ≥ 3 months 3rd: adult removed ≥ 6 months
Other policies	• Asset limit \$3,000 • Partial family cap (50 percent) • Two years transitional Medicaid • Child care assistance • Child support: \$100 disregard, full pass-through	• Asset limit \$1,000 • 100-hour rule and work history requirement for two-parent families • One-year transitional Medicaid • Child support: \$50 disregard, \$50 maximum pass-through

Source: Bloom et al. (2002).

to much recent discussion among policymakers and researchers, our results suggest the possibility that Connecticut's welfare reform reduced income for a nontrivial share of the income distribution after time limits took effect. Fourth, we find that the essential features of our empirical findings could not have been revealed using mean impact analysis on typically defined subgroups: the intragroup variation in QTE greatly exceeds the intergroup variation in mean impacts.

The remainder of the paper is organized as follows. In Section I, we provide an overview of the Jobs First program and its predicted effects. We then discuss our data in Section II. In Section III, we present empirical evidence that strongly suggests the time limit was an important program feature, and we present mean treatment effects in Section IV. Our main QTE results appear in Section V. We discuss extensions and sensitivity tests in Section VI, and we conclude in Section VII.

I. The Jobs First Program and Its Economic Implications

Below we compare the earnings, transfer, and income distributions between a randomly assigned treatment group, whose members face the Jobs First eligibility and program rules, and a randomly assigned control group, whose members face the AFDC eligibility and program rules. We begin by outlining the two programs and use labor supply theory to generate predictions about earnings, transfers, and income under Jobs First compared to AFDC.

Table 1 summarizes the major features of Connecticut's Jobs First waiver program and the existing AFDC program. The Jobs First waiver contained each of the key elements in PRWORA: time limits, work requirements, and financial sanctions. Jobs First's earnings disregard policy is quite simple: every dollar of earnings below the federal poverty line (FPL) is disregarded for purposes of benefit determination. This leads to an implicit tax rate of 0 percent for all earnings up to the poverty line, which is a very generous policy by comparison to AFDC's. The statutory AFDC policy disregarded the first \$120 of monthly earnings during a woman's first 12 months on aid, and \$90 thereafter. In the first four months, benefits were reduced by two dollars for every three dollars earned, and starting with the fifth month on aid, benefits were reduced dollar for dollar, so that the long-run statutory implicit tax rate on earnings above the disregard was 100 percent.[3]

[3] In practice, AFDC effective tax rates were less than the 100-percent statutory rate. First, there were work expense and

As shown in Table 1, the Jobs First time limit is 21 months, which is currently the shortest in the United States (Office of Family Assistance, 2003, Table 12:10). By contrast, there were no time limits in the AFDC program. In addition, work requirements and financial sanctions were strengthened in the Jobs First program relative to AFDC. For example, the Jobs First work requirements moved away from general education and training, focusing instead on "work first" training programs. Further, Jobs First exempts from work requirements only women with children under the age of one, and financial sanctions are supposed to be levied on parents who do not comply with work requirements. While Jobs First's sanctions are more stringent than AFDC's, the available evidence suggests that they were rarely used. For more information on these and other features of the Jobs First program, see our earlier working paper (Bitler et al., 2003b) and MDRC's final report on the Jobs First evaluation (Diana Adams-Ciardullo et al., 2002, henceforth the "final report").

Basic labor supply theory makes strong and heterogeneous predictions concerning welfare reforms like those in Jobs First. In the rest of this section, we discuss the economic impacts of Jobs First on the earnings, transfers, and income distributions. We focus on earnings disregards and time limits, since they are the salient features for examining heterogeneous treatment effects.

A. *Economic Impacts of Earnings Disregards*

To begin, Figure 1 shows a stylized budget constraint in income-leisure space before and after Jobs First. The AFDC program is represented by line segment *AB* while Jobs First is represented by *AF*. The Jobs First program dramatically affects the budget constraint faced by welfare recipients—lowering the benefit reduction rate to 0 percent and raising the breakeven earnings level to the FPL.[4] The effective AFDC benefit reduction rate in this figure is below the statutory long-run rate of 100 percent (see footnote 3 for a discussion).

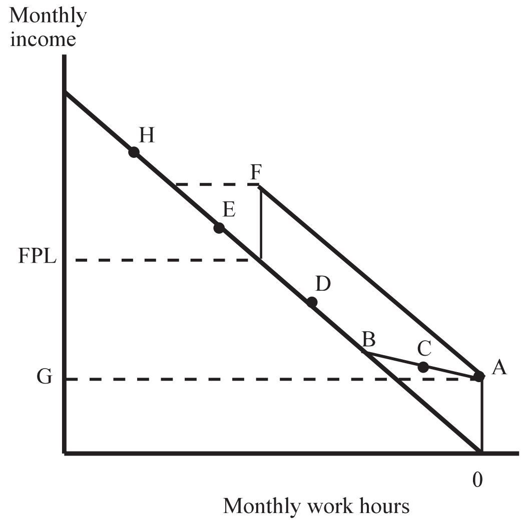

FIGURE 1. STYLIZED CONNECTICUT BUDGET CONSTRAINT UNDER AFDC AND JOBS FIRST

What is the impact of this transformation of the on-welfare budget segment from AFDC's *AB* to Jobs First's *AF*? To begin, we make the usual static labor supply model assumptions: the woman can freely choose hours of work at the given offered wage, and offered wages are constant. In particular, we ignore any human capital, search-theoretic, or related issues. We

child care disregards. Second, AFDC eligibility redetermination occurred less frequently than monthly, so there could be a lag between the month when an AFDC participant earned income and the date when benefits were reduced. Third, the Earned Income Tax Credit (EITC) provides a 40-percent wage subsidy in its phase-in region, which generally ended above Connecticut's maximum benefit level. (The EITC is available to both experimental groups in our data, so it raised the net wage above its before-tax level for both groups.) In Bitler et al. (2003b), we present local nonparametric regressions of transfer payments on earnings and find that the control group members receiving AFDC in our sample faced an effective benefit reduction rate of about one-third, similar to earlier studies of the national caseload in Terra McKinnish et al. (1999) and Thomas Fraker et al. (1985). Also, statutory rules for both AFDC and Jobs First tax away nonlabor income other than child support dollar for dollar; we discuss child support interactions in Section VIC.

[4] Under AFDC rules, eligibility for AFDC conferred categorical eligibility for food stamps, with a 30-percent benefit reduction rate applied to non–food stamps income. Under Jobs First, food stamps rules mirror those for cash assistance: food stamps benefits are determined after disregarding all earnings up to the poverty line (though this food stamps disregard expansion operates only while a woman assigned to Jobs First is receiving cash welfare payments). However, losing eligibility for welfare benefits under Jobs First assignment (e.g., through time limits) need not eliminate food stamps eligibility, since one could still satisfy the food stamps need standard.

TABLE 2—PRE–TIME LIMIT PREDICTED EFFECTS OF JOBS FIRST ASSIGNMENT, BY OPTIMAL CHOICE GIVEN AFDC ASSIGNMENT

Location if assigned to AFDC	Compared to this point, does Jobs First assignment change: After-tax wage?	Nonlabor income?	Location on Jobs First budget set	Effect on distribution of: Hours/earnings	Transfers	Income
A	Yes	No	*A*	0	0	0
	Yes	No	On *AF*, left of *A*	+	0	+
C	Yes	No	On *AF*, left of *C*	+	+	+
D	No	Yes	On *AF*, right of *D*	−	+	+
E	No	Yes	On *AF*, left of *A*	−	+	+
H	No	Yes	On *AF*, left of *A*	−	+	−
	No	No	*H*	0	0	0

Notes: Table contains predictions of static labor supply model for women facing AFDC and counterfactual Jobs First disregard rules (assuming all other rules are the same). Points are those labeled in Figure 1. There are two predictions for women at points A and H depending on those women's preferences.

also assume that there is no time limit. Later we relax these assumptions.

Consider first the case in which an AFDC-assigned woman locates at point *A*, working zero hours and receiving the maximum benefit payment *G*. Depending on the woman's preferences (e.g., the steepness of her indifference curves), assignment to Jobs First could lead to either of two outcomes. First, she might continue to work zero hours and receive the maximum benefit with no change in income. Second, she might enter the labor market, moving from *A* to some point on *AF*; transfer income remains at the maximum benefit level, while total income rises. This labor supply prediction—together with others discussed below—is summarized in Table 2, which indicates whether Jobs First changes the after-tax wage (in this case, yes) and nonlabor income (in this case, no). Table 2 then indicates the predicted location on the Jobs First budget set and the expected impact of Jobs First assignment on earnings, transfers, and income.[5]

We next consider points such as *C*, where women work positive hours and receive welfare when they are assigned to AFDC. For such women, assignment to Jobs First has only a price effect: the benefit reduction rate is lower, but there is no change in nonlabor income at zero hours of work. As long as substitution effects dominate income effects when only the net wage changes, Jobs First will cause an increase in hours, earnings, transfers, and income.

Now imagine that a woman's preferences are such that she would not participate in welfare if assigned to AFDC, instead locating at a point like *D*. At this point, her earnings would be between the maximum benefit amount and the FPL. Assignment to Jobs First would make this woman income-eligible for welfare even if she did not change her behavior; this is the case of Orley Ashenfelter's (1983) "mechanical" induced eligibility effect leading to an increase in transfers. If we assume that both leisure and consumption are normal goods, then we will expect the increase in nonlabor income accompanying Jobs First assignment to reduce hours of work and increase total income. That is, we expect women who would locate at point *D* to move to a point on *AF* that is both right of and above *D*.

Next consider a woman who would locate at a point like *E* if assigned to AFDC. At *E*, earnings are between the poverty line and the sum of the maximum benefit and the poverty line. Such points are clearly dominated under Jobs First assignment: the woman can increase income by reducing hours of work and claiming welfare (an example of Ashenfelter's behavioral induced eligibility effect). If both leisure and consumption are normal goods, we expect this woman to locate on *AF* at a point higher than *E*, so that hours worked decrease, while transfers and income both increase.

[5] Note that labor supply theory makes predictions about hours worked. Assuming no change in offered wages, this implies a prediction about earnings. Thus the table includes a single prediction for hours/earnings, which is important, since we observe earnings but not hours in our data.

Lastly, consider a woman who under AFDC assignment would locate at points like *H*, where earnings exceed the sum of the poverty line and the maximum benefit (above the notch). Depending on her preferences, Jobs First assignment will be associated with either of two possible outcomes. First, the woman might reduce hours of work so that her earnings fall to or below the poverty line; transfers increase and total income decreases. Reduced income in this case is compensated for by reduced disutility from labor; this is another example of Ashenfelter's (1983) behavioral induced eligibility effect. Alan S. Blinder and Harvey S. Rosen (1985) discuss the positive and normative implications of notches in budget constraints in more detail. Second, Jobs First assignment might have no effect for such women: if disutility of labor were sufficiently low, reduced labor hours would not fully compensate for the income lost in moving from *H* to *AF*, so the woman would stay at point *H*.

It is worth noting that because of the nature of the experiment, any Jobs First–induced entry into welfare (such as that experienced by those at points *D* and *E*) must come from reentry or decreases in exits, rather than new entry of nonrecipients. We discuss this issue in more detail in the next section.

The set of points {*A*, *C*, *D*, *E*, *H*} exhausts all qualitatively possible earnings-hours combinations under AFDC assignment. Thus, we use the final columns of Table 2 to summarize the impact of Jobs First on earnings, transfers, and income. For some part of the bottom of the distribution, the Jobs First earnings effect will be zero. At the very top of the earnings distribution, Jobs First will also have no effect on earnings, since top earners will choose to participate in neither AFDC nor Jobs First. In between these extremes, we expect the Jobs First earnings distribution to be higher at lower earnings levels, primarily due to increased labor force participation under Jobs First. Income effects for newly mechanically eligible women will tend to mitigate this prediction; which effect dominates is an empirical question. Lastly, there will be a range of earnings toward the top of the distribution where Jobs First earnings are lower than AFDC earnings due to behavioral induced eligibility effects.[6]

Before time limits take effect, static labor supply theory makes very simple predictions concerning the transfer-payments distribution: no one's transfers will fall, while some women will receive an increase in welfare payments (from zero to the maximum benefit). The effects at the bottom of the transfer payments distribution will be zero under either program, since some women will not receive welfare under either program assignment. Combining the predictions for earnings and transfers, the Jobs First earnings disregard reform is expected to lead to increased income throughout the distribution, with two notable exceptions. First, at the bottom of the distribution, we expect no change in income. Second, at the top of the distribution, income may either fall or stay the same.

B. *Economic Impacts of Other Jobs First Policies*

Jobs First features a 21-month time limit. Once the time limit takes effect, some women will no longer be eligible for any welfare benefits. For women who would have left welfare by then when assigned to the AFDC group, the time limit's effect on welfare payments is zero. Once the time limit binds, however, assignment to Jobs First rather than AFDC simply eliminates all welfare eligibility—i.e., once the time limit binds, Jobs First assignment eliminates the segment *AB* from the AFDC budget set. This change will obviously reduce transfer payments for women who would receive welfare if assigned to the AFDC group (thus all behavioral induced eligibility effects will disappear for time-limited women). The time limit will also increase labor supply due to the fall in nonlabor income and the rise in the net wage; no one's labor supply should fall as a result of the time limit's imposition. Thus, in comparing Jobs First to AFDC, we expect that when the time limit binds it will reinforce predicted positive earnings effects while eliminating predicted negative earnings effects. One can show that,

[6] We are not the first to point out that changes in the earnings disregard can lead to heterogeneous impacts on labor supply. The AFDC literature makes this point when discussing changes in the benefit reduction rate (see Moffitt's 1992 review for a discussion), and it is also discussed with varying emphasis in the recent welfare reform literature. For a useful summary of different policies for changing earnings disregards in welfare reforms, see Blank et al. (2000).

for some women, the increase in earnings will outweigh the loss in transfers, while the opposite is true for others; holding offered wages constant, we will generally expect increases in income to occur higher in the income distribution than decreases in income. Of course, the overall post–time limit impact of Jobs First on our three distributions will be a mix of the impacts for women bound by the time limit as well as those not bound.

With forward-looking behavior, the time limit may also have effects on women who have not yet exhausted eligibility. For now, we ignore this sort of behavior and focus on the results in the context of the change in earnings disregards. Later, in Section VI, we discuss the implications of these competing theories and present the available evidence.

Jobs First also brought a number of other reforms, including increased job search assistance, work requirements, sanctions for noncompliance, a more generous child support disregard and full child support pass-through, more generous asset limits, child care and medical insurance expansions, and family caps. With the exception of those related to child support, these changes are less important in the current context because they all lead to the prediction that labor supply should rise and welfare payments should fall. Jobs First's child support changes are potentially more significant for two reasons. First, these changes increase the amount of nonlabor income a woman may receive by increasing the child support disregard from \$50 to \$100, a change that should reduce labor supply. Holding constant the level of child support paid by fathers, however, the policy change can never increase disposable income by more than \$50 per month, an amount that we would not expect to affect behavior significantly. Second, for women receiving more than \$100 in monthly child support, Jobs First changed the distribution of funds across the child support and benefit checks. This change leads to a data asymmetry, which we discuss further in Section VIC.

II. Data

Under federal law, states were required to conduct formal evaluations when they implemented AFDC waivers. Connecticut fulfilled this requirement by hiring MDRC to conduct a random-assignment study of Jobs First. We use data made available by MDRC to outside researchers on completion of an application process. Random assignment in the Jobs First evaluation took place between January 1996 and February 1997, and data collection continued through the end of December 2000. The experimental sample includes cases that were ongoing (the recipient, or stock, sample) or opened (the applicant, or flow, sample) in the New Haven and Manchester welfare offices during the random assignment period. Assignment for recipients took place when they received an annual AFDC eligibility redetermination.

MDRC's evaluation and public-use samples include data on a total of 4,803 cases. Of these, 2,396 were assigned to Jobs First and 2,407 to AFDC. Rounded data on quarterly earnings and monthly welfare and food stamps income are available for most of the two years preceding program assignment and for at least four years after assignment. Further details are available in Appendix A. Demographic data—including information on number of children, educational attainment, age, race and ethnicity, marital status, and work history of the sample member—were collected at an interview prior to random assignment. During the evaluation period, Connecticut's nonexperimental caseload was moved to Jobs First; with only a few exceptions, only the experimental control group continued under the AFDC rules.

At this point it is important to recall that we are interested in the labor supply choices of women under counterfactual assignment to Jobs First and AFDC. All women in the experiment have applied for public assistance, and most are not working at the time of random assignment, so most begin at point *A* of Figure 1. We observe women for four years, however, and over time women leave welfare. For example, we find that about half of women in the AFDC control group have left welfare within two years after random assignment, which is similar to the pattern of welfare dynamics in the literature (Mary Jo Bane and David T. Ellwood, 1994). Differences in offered wages, fixed costs of work, and/or preferences will cause AFDC-assigned women to leave welfare at different rates and end up at different points on the counterfactual budget set. In thinking about Jobs First's impacts on counterfactual outcomes, we have in mind simply that at some point after random assignment, a woman assigned to the AFDC group may choose to locate at points like D, E, or H. It is important to note that any increase in welfare participation due to Jobs First assignment in such cases can be due either to increased reentry or (even more likely) to reduced exit—but it will not be due to increases in first-time welfare participation.[7]

The first column of Table 3 provides means for a national sample of AFDC recipients in 1996,[8] while the next two columns provide means for the same characteristics among women in our experimental data. The Jobs First experimental sample generally mirrors the characteristics of the national sample, with the exceptions that the experimental sample has substantially greater fractions of never-married and less-educated women compared to the national caseload.

The bottom part of the table reports statistics for the experimental treatment and control samples concerning pretreatment earnings, employment, and welfare use, as well as whether women came from the experimental recipients sample. The fourth column of the table reports unadjusted differences across the program groups. Overall, demographic characteristics are substantively similar across the experimental program groups. There are potentially important (if not large) exceptions, however: the Jobs First group had lower earnings, greater cash welfare use, larger families (not shown), and a greater share of the sample coming from the recipients sample than did the AFDC group. A test for joint significance of the 17 linearly independent differences in Table 3 yields a χ^2 test statistic of 24 (p = 0.12), so we cannot reject that program assignment was indeed random.[9]

Even though we cannot reject random assignment, one might worry about differences in pretreatment earnings and welfare variables. One approach to dealing with such unbalanced samples would be to include pretreatment variables as covariates in a quantile regression model (the analogue of adding them to a linear regression model). Instead, we take the more theoretically appropriate approach of using inverse-propensity score weighting; we discuss the weights and the logit model used to estimate them in Appendix B. The final column of Table 3 reports estimated differences after adjusting using inverse propensity score weighting. As statistical theory predicts, the differences are reduced to almost exactly zero. We use inverse-propensity score weights in the standard fashion for all estimators employed below. It is important to point out that the propensity score adjustment does not alter our qualitative conclusions, which hold whether we weight or not, and whether or not our propensity score model includes demographic controls. Unweighted results are available on request.

III. Empirical Evidence on the Time Limit

As stated above, Jobs First's 21-month time limit is currently the shortest in the United States. About 29 percent 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 (see the final report for discussion). Under certain circumstances, Jobs First caseworkers were empowered to provide both indefinite exemptions from the time limit and to provide six-month extensions. According to the final report, in the spring of 1998, 26 percent of the statewide (not just the experimental) caseload was exempt from the time limit. This number rose to

[7] In part to mitigate possible entry effects, the Jobs First program has "dual eligibility" rules. While the FPL is used to determine continuing eligibility for current *recipients,* successful *applicants* must have monthly earnings no greater than $90 plus the state welfare needs standard (which was $745 for a family of three in 1999), leading to a considerably more stringent earnings test for applicants than for recipients (whose earnings need only be below the poverty line, which was $1,138 for a family of three). This dual eligibility policy will tend to reduce the earnings level at which any actual entry effects occur, but it will not eliminate all entry incentives, nor does it mitigate the deterred-exit effects we discuss above. Since static labor supply analysis is qualitatively unaffected by the dual eligibility rule for applicants, we do not address it separately here.

[8] The estimates for the national caseload are constructed using March 1997 CPS data. The sample includes all women age 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.

[9] In prior versions of this paper, we reported highly significant values for this statistic. These values were erroneous, resulting from a now-fixed programming error.

TABLE 3—CHARACTERISTICS OF NATIONAL CASELOAD AND EXPERIMENTAL SAMPLE

	National caseload (CPS)	Experimental sample			
		Levels		Differences	
		Jobs First	AFDC	Unadjusted	Adjusted
Demographic characteristics					
White	0.405	0.362	0.348	0.014	0.001
Black	0.344	0.368	0.371	−0.003	−0.000
Hispanic	0.206	0.207	0.216	−0.009	−0.001
Never married	0.474	0.654	0.661	−0.007	−0.000
Div/wid/sep/living apart	0.316	0.332	0.327	0.005	0.000
HS dropout	0.399	0.350	0.334	0.017	−0.000
HS diploma/GED	0.358	0.583	0.604	−0.021	0.001
More than HS diploma	0.243	0.066	0.062	0.004	0.000
More than two children	0.280	0.235	0.214	0.021*	−0.000
Mother younger than 25	0.251	0.289	0.297	−0.007	−0.000
Mother age 25–34	0.436	0.410	0.418	−0.007	0.000
Mother older than 34	0.313	0.301	0.286	0.015	0.000
Recipient (stock) sample		0.624	0.593	0.031**	−0.001
Average quarterly pretreatment values					
Earnings		679	786	−107***	−1
		(1,304)	(1,545)	(41)	(32)
Cash welfare		891	835	56**	−1
		(806)	(785)	(23)	(2)
Food stamps		352	339	13	0
		(320)	(304)	(9)	(1)
Fraction of pretreatment quarters with					
Any earnings		0.322	0.351	−0.029***	0.000
		(0.363)	(0.372)	(0.011)	(0.001)
Any cash welfare		0.573	0.544	0.029**	−0.001
		(0.452)	(0.450)	(0.013)	(0.001)
Any food stamps		0.607	0.598	0.009	0.000
		(0.438)	(0.433)	(0.013)	(0.001)

Notes: Numbers in parentheses are standard deviations for columns 2 and 3; those for columns 4 and 5 are standard errors. For all but column 5, these statistics are estimated conventionally; for column 5, we compute standard errors using 1,000 nonparametric bootstrap replications. ***, **, and * indicate statistical significance at the 1-percent, 5-percent, and 10-percent levels, respectively (significance indicators provided only for difference estimates). National caseload statistics were constructed using all females age 16–54 in the 1997 March CPS who had an own child in the household and whose family was reported to have positive AFDC income for calendar year 1996. All national caseload statistics are computed using March supplementary weights. Standard deviations omitted for binary variables. For earnings, eight quarters of pretreatment data are used. For cash welfare and food stamps, only seven quarters are available for all observations. Baseline data on a small number of observations for some variables are missing.

49 percent by March 2001; the increase appearsto largely reflect progressive exits from the caseload by more able (and time-limited) recipients. A woman could receive an indefinite exemption because of mental incapacitation, responsibility to care for a disabled relative, having a child aged younger than one, and being deemed unemployable due to limited work history and human capital. Extensions were granted to a nonexempt woman if her family income was below the applicable maximum benefit payment and she had made a good-faith effort to find and retain employment. If no good-faith determination was made, then an extension was still possible if "there were circumstances beyond the recipient's control that prevent[ed] her from working" (final report, p. 63).

In light of these statistics, it is critical to show that the time limit policy has de facto relevance. We do so using Figure 2. The solid line in the figure plots the treatment effect due to Jobs First on the first-spell survival function. This series is calculated as the Jobs First group's first-spell survival function minus the AFDC group's survival function, with the latter plotted as the

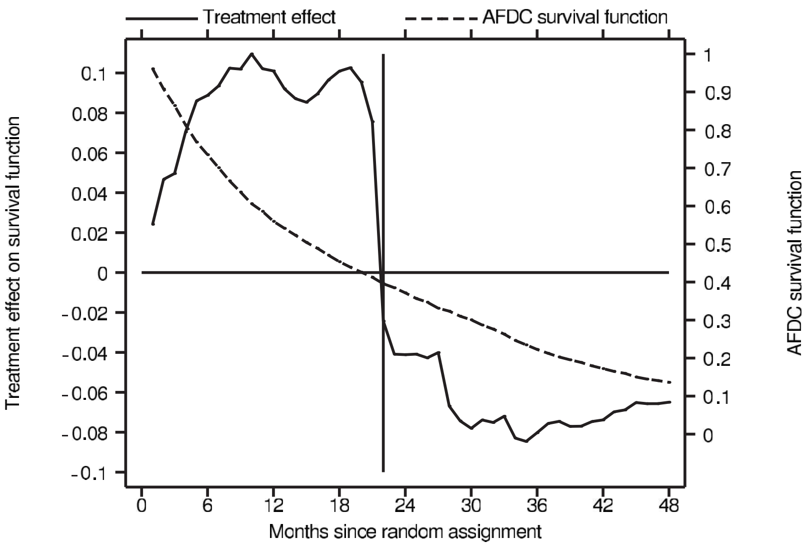

FIGURE 2. FIRST-SPELL MONTHLY SURVIVAL FUNCTION: AFDC GROUP AND TREATMENT EFFECT

Notes: All statistics computed using inverse propensity-score weighting. See text for more details.

smooth, dashed line in the figure.[10] Figure 2 has five key features. First, the treatment effect of Jobs First on first-spell survival is actually positive throughout the pre–time limit period, reflecting the increased generosity of the program before time limits take effect. Second, there is a sharp drop of 10 percentage points in the survival treatment effect between months 21 and 22: exactly the point when time limits can first bind. Third, the treatment effect on welfare participation is negative after this point. Fourth, the time limit was not binding for everyone. At month 22 the control group survival rate was 40 percent, with the month-22 treatment effect being -0.024. This is, of course, just another way of saying that exemptions and extensions were provided, as we knew. Fifth, there are (smaller) sharp drops at the six-month intervals when extensions expire. Overall, Figure 2 provides compelling evidence that the time-limit policy was binding for a substantial number of women. This fact is the important one for our purposes.

IV. Mean Treatment Effects

The first column of Table 4 reports estimated mean levels among the Jobs First group for several variables, over the entire 16-quarter posttreatment period. The second column provides means for the AFDC group over the same period, and the third column provides the resulting mean impacts. The first three rows contain average quarterly values of total income (defined as the sum of earnings and total transfers), earnings, and total transfers (defined as AFDC or Jobs First cash payments plus food stamps). These results show that over the four years following random assignment, the impact of Jobs First on average total income was $136 (about 5 percent compared to estimated AFDC baseline quarterly income of $2,609). About two-thirds of this impact is due to an insignificant increase in earnings, with the remainder

[10] We label women as being in their first spell at the beginning of this period if they have cash welfare income in either the month of random assignment or the following month. For those in the recipient sample, some spells will have ended coincidentally in the month of random assignment; for those in the applicant sample, not all applications will be accepted. For these reasons, only 85 percent of AFDC-assigned women and 88 percent of Jobs First–assigned women are in a first spell at the beginning of the analysis period. Note that this first-spell definition does not match the usual one, since we include all ongoing spells, whether or not they are left-censored. Typically in such pictures, the time origin is set to be the beginning of a fresh spell, whereas our time origin is the time of experimental assignment, which may or may not coincide with the beginning of a welfare spell.

TABLE 4—MEAN OUTCOMES AND IMPACTS

	All quarters			Quarters 1–7			Quarters 8–16		
	Jobs First	AFDC	Adjusted difference	Jobs First	AFDC	Adjusted difference	Jobs First	AFDC	Adjusted difference
Average quarterly level									
Income	2,745	2,609	136**	2,744	2,450	294***	2,748	2,733	14
	(35)	(57)	(64)	(31)	(48)	(53)	(44)	(67)	(78)
Earnings	1,658	1,561	97	1,195	1,113	82	2,020	1,908	112
	(35)	(58)	(64)	(29)	(49)	(52)	(45)	(68)	(78)
Transfers	1,088	1,048	40**	1,550	1,337	212***	728	825	−98***
	(15)	(16)	(20)	(17)	(17)	(22)	(17)	(18)	(23)
Fraction of quarters with									
Any income	0.852	0.857	−0.005	0.908	0.906	0.002	0.809	0.820	−0.010
	(0.005)	(0.005)	(0.007)	(0.005)	(0.005)	(0.006)	(0.007)	(0.006)	(0.009)
Any earnings	0.561	0.490	0.071***	0.519	0.442	0.077***	0.593	0.527	0.066***
	(0.007)	(0.007)	(0.009)	(0.007)	(0.007)	(0.009)	(0.008)	(0.008)	(0.011)
Any transfers	0.626	0.622	0.004	0.794	0.756	0.038***	0.496	0.519	−0.023**
	(0.007)	(0.007)	(0.009)	(0.006)	(0.007)	(0.009)	(0.008)	(0.009)	(0.011)
N	2,381	2,392	4,773	2,396	2,407	4,803	2,381	2,392	4,773

Notes: Standard errors in parentheses calculated using 1,000 nonparametric bootstrap replications. ***, **, and * indicate statistical significance at the 1-percent, 5-percent, and 10-percent levels, respectively (significance indicators provided only for impact estimates). All statistics computed using inverse propensity-score weighting.

due to a significant increase in transfers of $40, an effect of about 4 percent.

The bottom three rows provide means and impacts for binary variables indicating the fraction of quarters for which the person had positive levels of income, earnings, and transfers in the full 16-quarter period. For example, the value of 0.852 for "any income" means that among women assigned to Jobs First, 85.2 percent of all person-quarters had a positive value for at least one of unemployment insurance (UI) earnings, cash assistance, or food stamps.[11] The results show that the probability of having any earnings was 7.1 percentage points greater among the Jobs First group than the control group, an effect of 14 percent relative to the control group baseline. The probability of having any income or any transfers is essentially identical across treatment status over the full 16-quarter period.[12]

Both theory and the evidence above on the time limit suggest that in the first 21 months after random assignment—before time limits bind for anyone—effects induced by Jobs First are very different from effects during the final 27 months. Thus, we separately estimate mean treatment effects for the pre– and post–time limit periods. The second set of columns concerns the first seven quarters of data, while the third set concerns the last nine quarters. The results suggest that average earnings increased 7 percent in the pre–time limit period and 6 percent in the post–time limit period; in each case this effect is insignificant, though Jobs First significantly increases the fraction of person-quarters with any earnings in each period. Mean impacts for transfers are starkly different in the early and later periods. During the first seven quarters, Jobs First members received $212—or 16 percent—more in transfers than did control group women. During the later period, Jobs First members received $98—or 12 percent—less in transfers. The same pattern is clear for the fraction of person-quarters with positive transfers.

The net result of these changes in earnings and transfers is that Jobs First increased mean total income significantly—in both economic and statistical terms—in the pre–time limit pe-

[11] This also means that about 15 percent of person-quarters had no value in any quarter for any of these variables, which could mean that 15 percent of persons never have any income, that everyone has positive income for all but 15 percent of quarters, or something in between. We return to this issue below.

[12] The share having any earnings can increase even while the share having any income does not because women caused to work by Jobs First assignment would have had welfare income if assigned to AFDC.

riod. Nearly three quarters of this increase came from increased transfer income rather than earnings. By contrast, mean income in the post–time limit period was virtually identical across treatment status, the result of nearly equal increases in mean earnings and reductions in mean transfers.

V. Quantile Treatment Effects

Before presenting our QTE estimates, it will be helpful to review briefly quantiles and QTE. For any variable Y having *cdf* $F(y) \equiv Pr[Y \leq y]$, the q^{th} quantile of F is defined as the smallest value y_q such that $F(y_q) = q$. If we consider two distributions F_1 and F_0, we may define the QTE as $\Delta_q = y_q(1) - y_q(0)$, where $y_q(t)$ is the q^{th} quantile of distribution F_t. This treatment effect may be seen to equal the horizontal distance between the graphs of F_1 and F_0 at probability value q; equivalently, it is the vertical distance between the graphs of the inverse *cdf*s. (Inverse *cdf* plots for all the variables we consider below are available on request from the authors, as are tables of the sample quantiles themselves.) The QTE estimates we report below are constructed in exactly this fashion: to estimate the QTE at the q^{th} quantile, we calculate the q^{th} quantile of the given Jobs First distribution and then subtract the q^{th} quantile of the given AFDC distribution. As a simple example, estimating the QTE at the 0.50 quantile simply involves taking the sample median for the treatment group and subtracting the sample median for the control group; we briefly discuss some technical details in Appendix B. Appendix Table 1 reports the deciles of the Jobs First and AFDC group distributions for each outcome variable and time period we consider below; readers may thus calculate the QTE at these deciles. This table is also useful for assessing the magnitude of the estimated QTE relative to the control group's baseline.

At this point, we want to emphasize that QTE do not necessarily identify the impact of treatment for given people. For example, if Jobs First causes rank reversals in the earnings distribution, then knowing the difference of medians in the two distributions is not enough to calculate the Jobs First treatment effect for a person who would have median earnings when assigned to AFDC. It is easy, however, to see that if any of the QTE is negative (positive), then the treatment effect must also be negative (positive) for some nondegenerate interval of the counterfactual AFDC earnings distribution. We also note that, like QTE estimates, classical social welfare function analysis would require only the empirical distributions of the two program groups. We discuss these and related issues in more detail in Bitler et al. (2003b).[13]

A. *QTE for Earnings*

We now turn to our main results: QTE for 97 centiles in graphical form.[14] Since we use the person-quarter as the unit of analysis, there are $7 \times 4{,}803 = 33{,}621$ observations for the first seven quarters. For the last nine quarters, there are $9 \times 4{,}773 = 42{,}957$ observations. (As discussed in Appendix A, we lack quarter-16 data on 30 experimental participants.)[15] To construct confidence intervals for the quantile treatment effects and test hypotheses discussed below, we use 1,000 bootstrap repetitions to estimate the sampling distribution of the estimated QTE, using the fifth and ninety-fifth percentiles of these distributions to construct equal-tailed estimated 90-percent confidence intervals; we provide details in Appendix B.

We plot the earnings QTE (as a solid line) for the first seven quarters after assignment in Figure 3. Dotted lines provide the bounds of 90-percent confidence intervals. For comparison purposes, the mean treatment effect is plotted as a horizontal (dashed) line, and the 0-line is provided for reference. This figure shows that for quarterly earnings in the pre–time limit period, the QTE are identically zero for almost all quantiles below the median. This result occurs because quarterly earnings are identically 0 for 48 percent of person-quarters in the Jobs First

[13] Heckman et al. (1997) provide a more general discussion of treatment effect heterogeneity and associated normative analysis issues. For a discussion of the potential outcomes framework undergirding our work, see for example that paper and Guido W. Imbens and Joshua D. Angrist (1994).

[14] We computed QTE at quantiles 98–99 but do not include them in the figures below because their sampling variability is very great. We do not have the same problem at the bottom of the distributions because they are all bounded below by zero.

[15] In Bitler et al. (2003b) we also present QTE for earnings, transfers, and income averaged over the first seven and last nine quarters. The results are qualitatively similar to those presented here.

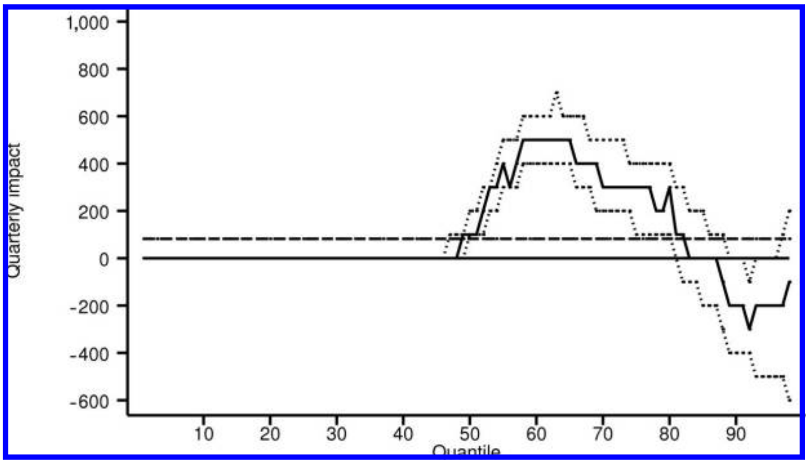

FIGURE 3. QUANTILE TREATMENT EFFECTS ON THE DISTRIBUTION OF EARNINGS, QUARTERS 1–7

Notes: Solid line is QTE; dotted lines provide bootstrapped 90-percent confidence intervals; dashed line is mean impact; all statistics computed using inverse propensity-score weighting. See text for more details.

group over the first seven quarters and 55 percent of corresponding AFDC group person-quarters. For quantiles 49–82, Jobs First group earnings are greater than control group earnings, yielding positive QTE estimates. Between quantiles 83 and 87, earnings are again equal (though non-zero). Finally, for quantiles 88–97, AFDC group earnings exceed Jobs First group earnings, yielding negative QTE estimates. The only quantile having a statistically significant QTE based on a two-sided test is the ninety-second—for all other quantiles between 89 and 96, the two-sided QTE confidence intervals include zero in the confidence interval. On the other hand, one-sided tests yield *p*-values of 0.10 or lower for all QTE in the 90–95 quantile range.[16] These results are what basic labor supply theory, discussed above, predicts. That is, the QTE at the low end are zero, they rise, and then they eventually become negative (if imprecisely estimated). The negative effects at the top of the earnings distribution are particularly interesting given that they have typically not been found in other programs (e.g., Nada Eissa and Jeffrey B. Liebman's, 1996, study of the EITC).

The variation in Jobs First's impact across the quantiles of the distributions appears unmistakably significant, both statistically and substantively; these results suggest that the mean treatment effect is far from sufficient to characterize Jobs First's effects on earnings.[17]

Figure 4 plots the earnings QTE results in quarters 8–16, after the time limit takes effect for at least some women. For the first 76 quantiles, these results are broadly similar to those for the pre–time limit period (though they have a somewhat wider range and become positive slightly earlier). For quantiles 77–97, we again find negative treatment effects (with a few being zero), but none of them is individually signifi-

[16] To test whether these QTE estimates are jointly significantly negative, we carry out two sets of tests. Details are somewhat complicated, so we relegate them to Appendix B. Our basic conclusion, however, is that there is some marginal evidence that these QTE are jointly different from zero.

[17] Under the null of constant treatment effects, all QTE must equal the mean treatment effect. This null can be rejected decisively simply by noting the large fraction of the treatment group earnings distribution having zero earnings (Heckman et al., 1997, make a similar point regarding treatment effects of job training). We did conduct more formal tests for the null that the \$800(= \$500 − (−\$300)) range of the estimated QTE could have been generated under the null that all quantiles of the Jobs First distribution equal the mean treatment effect plus the corresponding quantiles of the AFDC distribution. These tests, which impose the null by using paired bootstrap sample draws from the AFDC group sample and then adding the mean treatment effect to each sample quantile in one of the pairs, soundly reject the equality of the QTEs.

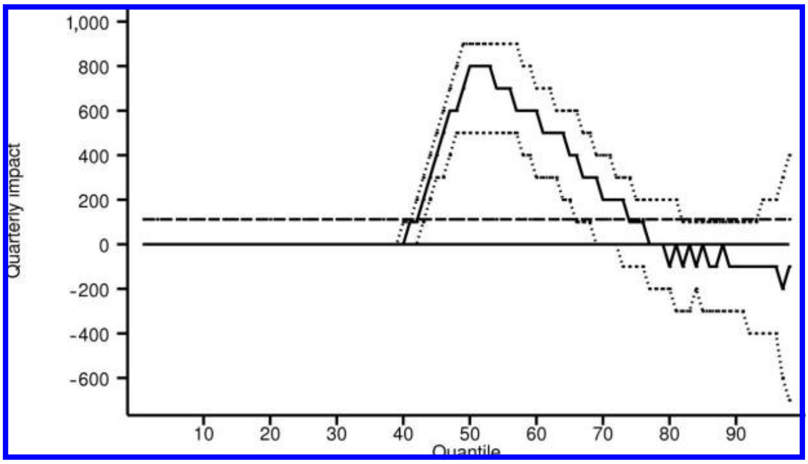

FIGURE 4. QUANTILE TREATMENT EFFECTS ON THE DISTRIBUTION OF EARNINGS, QUARTERS 8–16

Notes: Solid line is QTE; dotted lines provide bootstrapped 90-percent confidence intervals; dashed line is mean impact; all statistics computed using inverse propensity-score weighting. See text for more details.

cant. Again, these results match the predictions of basic labor supply theory—the QTE are zero at the low end, positive in the middle, and zero at the top of the distribution. Recall that static labor supply theory predicts zero effects at the top of the earnings distribution in the post–time limit period, given that, under AFDC, women with moderate- to high-earnings capacity are likely to already be off welfare by this time.

The expanded disregard can reduce earnings via reentry or nonexit only while women retain welfare eligibility. There are two sets of women who can be eligible for Jobs First welfare even after month 21: those who left welfare before month 21, and those who receive exemptions or extensions. Women in the first group are unlikely candidates for behavioral induced eligibility effects after the seventh quarter, given the fact that they have already left once, together with the more stringent earnings test for reentry (see footnote 7 above). Getting an extension or exemption generally requires having earnings below the maximum benefit level, which is typically substantially below the poverty line (the difference depends on family size). It seems particularly unlikely that the Jobs First notch would cause reentry or nonexit effects for these women. Thus, static labor supply theory predicts significant behavioral induced eligibility effects in the first seven—but not the last nine—quarters of the Jobs First experiment. This pattern is exactly the one we see. Consequently, we suspect that the reduction in earnings at the top of the distribution caused by Jobs First in Q1-Q7 is most likely due to behavioral induced eligibility effects of the disregard expansion.

B. *QTE for Transfers*

Figure 5 presents results for transfer income in the first seven quarters, and Figure 6 presents results for the last nine quarters. The most notable feature of these results is the radical difference in the treatment effects of Jobs First across the pre– and post–time limit period. In the first seven quarters, the QTE are identically 0 for the bottom 20 quantiles, reflecting the fact that for the bottom fifth of the distribution, both the treatment and control group have zero transfer income. For all quantiles except two above the twentieth, transfer income in the pre–time limit period is greater among Jobs First women than among AFDC women. This finding greatly extends the result for mean treatment effects presented in Section IV. Moreover, the range of QTE in this period is very large, with the largest QTE reaching $700. As a basis of comparison, this is nearly one-third of the maximum quarterly value of Connecticut's combined AFDC–food stamps payment for a family of three. Thus, in the pre–time limit period, Jobs First

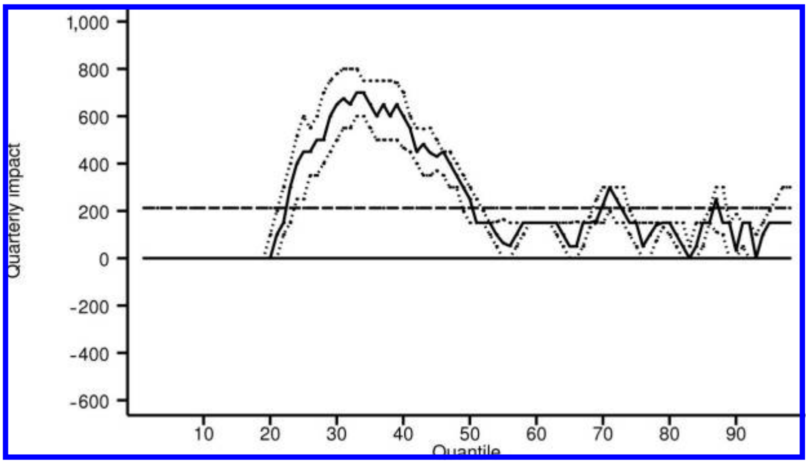

FIGURE 5. QUANTILE TREATMENT EFFECTS ON THE DISTRIBUTION OF TRANSFERS, QUARTERS 1–7

Notes: Solid line is QTE; dotted lines provide bootstrapped 90-percent confidence intervals; dashed line is mean impact; all statistics computed using inverse propensity-score weighting. See text for more details.

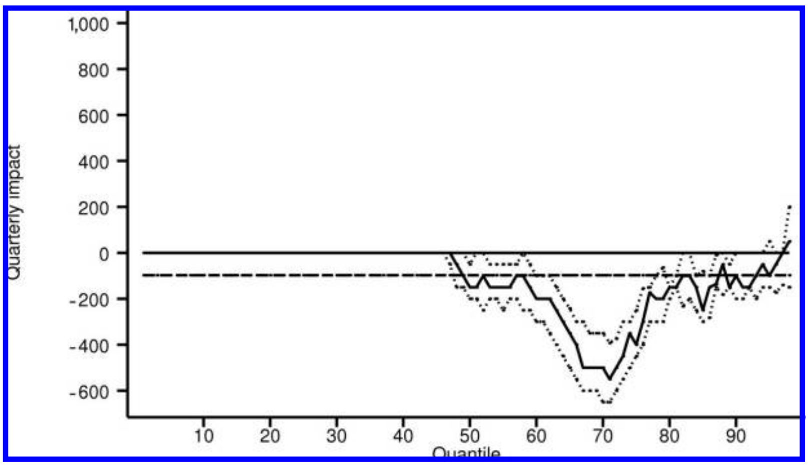

FIGURE 6. QUANTILE TREATMENT EFFECTS ON THE DISTRIBUTION OF TRANSFERS, QUARTERS 8–16

Notes: Solid line is QTE; dotted lines provide bootstrapped 90-percent confidence intervals; dashed line is mean impact; all statistics computed using inverse propensity-score weighting. See text for more details.

clearly is associated with a substantial upward shift in transfers over most of the distribution, as would be expected either from the simple mechanical effect of a more generous benefit schedule or from behavioral responses. Furthermore, the pattern of the QTE is consistent with theoretical predictions: little or no increase at the very top of the transfer distribution (which is both theoretically and empirically likely to be the bottom of the earnings distribution) or the very bottom (where no one participates) and increases in transfers everywhere in between.

The graph for quarters 8–16 is much different. For the lowest 47 quantiles, the Jobs First and AFDC transfer distributions are equal, with both showing zero transfer income at all these quantiles. However, at all quantiles between 48 and 96, the Jobs First group receives less trans-

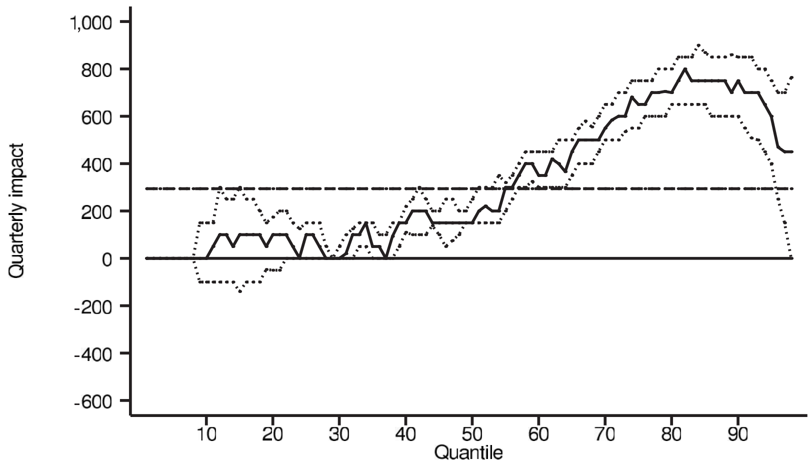

FIGURE 7. QUANTILE TREATMENT EFFECTS ON THE DISTRIBUTION OF INCOME, QUARTERS 1–7

Notes: Solid line is QTE; dotted lines provide bootstrapped 90-percent confidence intervals; dashed line is mean impact; all statistics computed using inverse propensity-score weighting. See text for more details.

fer income. The size of the reductions in transfer income can be quite large: the largest quarterly reduction is $550, and the reduction is at least $300 for all quantiles from 64 to 76. Results not reported here show that most of this reduction is due to the smaller fraction of Jobs First than AFDC women who receive any cash assistance in quarters 8–16 (the difference is 8 percent, compared to a 2-percent-larger fraction of Jobs First than AFDC women who had income from at least one of food stamps and cash welfare). When we estimate QTE results for cash assistance ignoring food stamps and including only those person-quarters having positive cash assistance, the QTE estimates are actually almost all positive. This result reflects the more generous Jobs First disregard, given eligibility. Thus, the negative QTE results for transfer payments in Figure 6 are primarily driven by reductions in the rate of cash assistance, which shifts the entire transfer *cdf* leftward, so that the inverse *cdf* shifts downward.

C. *QTE for Total Income*

We plot QTE results for total measurable income (earnings plus food stamps plus the amount of cash assistance sent by the state to welfare recipients) in the pre–time limit period in Figure 7. These results again suggest a large degree of treatment effect heterogeneity: they range from 0 for the bottom 10 quantiles—where total income is 0 in both groups—to $800 at the top of the range. The mean treatment effect for this period is $294, so again the range of quantile treatment effects is large compared to the mean treatment effect. It would be interesting to decompose QTE for income into a function of the marginal QTE for earnings and for transfers. However, there need not be any particular relationship between QTE for total income and QTE for its components. Without strong assumptions (e.g., rank preservation), it is impossible to draw general conclusions about the relationship between QTE for the various distributions.

Figure 8 plots QTE results for the post–time limit period. The figure clearly shows that Jobs First affects the distribution of total income, in stark contrast to the trivial mean treatment effect of $14. QTE estimates for total income are zero for the first 18 quantiles and are actually negative for the next 24 quantiles; the largest estimated reduction in quarterly total income is $300.

Before the adoption of PRWORA, many welfare advocates expressed great concern that welfare reform would harm large numbers of (actual or potential) welfare recipients. Yet a common conclusion in the welfare reform literature is that few if any welfare recipients have been harmed. Given relatively short lifetime

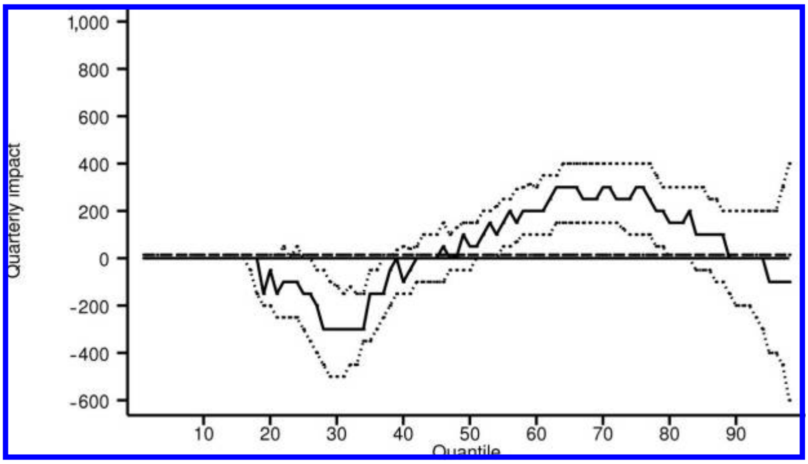

FIGURE 8. QUANTILE TREATMENT EFFECTS ON THE DISTRIBUTION OF INCOME, QUARTERS 8–16

Notes: Solid line is QTE; dotted lines provide bootstrapped 90-percent confidence intervals; dashed line is mean impact; all statistics computed using inverse propensity-score weighting. See text for more details.

time limits, our results for quarters 8–16 are more appropriate than the pre–time limit results for addressing this issue. While our approach captures the effects of welfare reform on the distribution of income, we cannot generally make a statement about how any individual person's income is changed. We find it notable, however, that when we use individual significance tests, ten of our total-income QTE estimates for quarters 8–16 are significantly negative, while 25 are significantly positive. To test the significance of these findings, we use a bootstrap test of the null hypothesis that Jobs First causes no negative total income QTE, as well as the null that Jobs First causes no positive QTE (see Appendix B for details on how to use Abadie's, 2002, method to impose the null and on the test statistics themselves). A bootstrap test rejects the null hypothesis that all QTE for centiles 19–38 are nonnegative ($p = 0.075$); we also reject the null that the largest negative QTE in that range (−\$300) is zero ($p = 0.022$). A test for whether centiles 48–88 are positive rejects ($p = 0.065$) the null that all 41 of these quantiles are nonpositive; lastly, we easily reject the null that the largest positive QTE in that range (\$300) is zero ($p = 0.009$).

We draw two conclusions from this analysis. First, once time limits take effect, there are definitely negative effects on some women, and positive effects on others. Second, the evidence of positive effects occurs over a larger range of the distribution. Even more informative, the negative income QTE are concentrated at the lower end, with the positive ones concentrated in the upper half. As usual, when there are both winners and losers, resolving these opposing results would require the use of some normative metric, which is beyond the scope of this paper.[18]

VI. Extensions and Sensitivity Tests

A. *Banking and Queuing*

In the section above, we show that changes in the earnings and transfer distribution are very much in line with the labor supply incentives embodied in the Jobs First expansion in the earnings disregard. Here we explore the possible role played by time limits in explaining the findings above.

[18] Without further assumptions, the possibility of rank reversals prevents us from being more specific about who the winners and losers are. If rank reversals occur in such a way as to minimize the number of losers, however, then the losses of these losers will be particularly large. Equity concerns are thus not necessarily mitigated in such cases. In Bitler et al. (2003b), we provide a normative analysis using a class of traditional social welfare functions, with the functions' parameters allowed to vary; such an approach does not depend on ranks.

In the static labor supply model, the presence of time limits does not affect labor supply and welfare participation until a woman reaches the time limit. With forward-looking behavior, however, time limits may affect labor supply before the time limit hits through "banking" and "queuing" motives.

Banking effects may lead a woman to conserve her eligibility by reducing welfare use and increasing labor supply even before the time limit binds, as discussed in Grogger and Michalopoulos (2003). What are the implications of this behavior for the distribution of earnings and transfers? First, the banking effect will cause more women to work under Jobs First, increasing earnings relative to AFDC. Second, the desire to leave welfare earlier should reduce both search durations and accepted wages (through lower reservation wages for working); this effect should reduce earnings among Jobs First women relative to AFDC. Therefore, like earnings disregards, banking may lead to either increases or decreases in earnings. Banking provides a qualitatively different prediction for transfers, however: transfers should fall prior to time limits, whereas static labor supply theory predicts an increase due to the expanded earnings disregard.

Further, with scarce jobs and a need to search for work, one might expect the time limit to cause women to accept lower offered wages (or cease human capital investment earlier than planned) even in the absence of any banking effect. An effect like this one could occur if women are worried about having to wait in a "job queue" after the time limit hits.[19] Such a queuing effect would have the same qualitative effects on the earnings distribution as the banking effect. It would not necessarily change the transfers distribution, however, since employed Jobs First women might choose to stay on welfare until time limits hit.

Consistent with the banking and queuing effects, data from MDRC's three-year follow-up survey (given to a subset of the full sample) suggest that among employed women, wages throughout much of the top half of the wage distribution are lower for women in the Jobs First group compared to those in the AFDC group. Unless wage growth is correspondingly greater, however, among those who take lower starting wages, we would expect negative banking and queuing effects on the earnings distribution to persist throughout the study period, even after women leave welfare. Yet in the results reported in Section V, we see no significant negative earnings QTE in the period after time limits hit. Thus, behavioral induced eligibility effects appear more consistent with the observed pattern of negative QTE at the top of the earnings distribution than do banking or queuing effects.

We can offer additional evidence to distinguish the behavioral induced eligibility effect and the banking effect. The banking effect implies not only that women should enter employment at lower wages, but also that they should exit welfare at lower wages. If lower reservation wages for exiting welfare were the only cause of reduced earnings at the top, then welfare participation rates at higher earnings levels should be lower among Jobs First than among AFDC women. To examine this hypothesis, we first sort person-quarter observations on earnings into ten bins corresponding to deciles of the AFDC group's earnings distribution; we do this separately for the first seven and last nine quarters. We then define an indicator variable equal to one when a woman has cash welfare income each month of a quarter, and zero otherwise. The banking effect suggests that in the pre–time limit period, the fraction with welfare income each month should be lower in the Jobs First group among women with relatively high quarterly earnings. We find the opposite to be true: for the AFDC group, 23 percent of women in decile 9 and 10 percent of women in decile 10 have welfare income each month of the quarter; for the Jobs First group, the corresponding numbers are 62 percent and 26 percent.

Further evidence on the queuing effect is more difficult to provide. Like the behavioral eligibility and banking effects, the queuing effect implies that earnings should fall at the top of the distribution. Unlike the banking effect, it does not imply that welfare participation should also fall. The only prediction that would allow us to distinguish between queuing and behavioral eligibility involves "bunching at the kinks" of the budget set. In particular, the large notch in the Jobs First budget set should lead to a mass point in the earnings distribution at the poverty

[19] We are grateful to an anonymous referee for suggesting this possibility.

line. We would therefore expect a spike in the density at that point, with a discontinuous drop occurring right above the poverty line. Of course, such bunching at the poverty line will hold only if women can perfectly choose their hours. If many women cannot adjust their hours, we would instead expect increased density over some earnings range below the poverty line and a discontinuous drop in the density at the poverty line.[20]

While our data are not ideal for this exercise, we explored whether there is bunching at the kinks using the method in Justin McCrary (2005).[21] Overall, we find no evidence of a discontinuous drop in the density at the poverty line, but we do see an increase in the density over a quarterly earnings range between the poverty line and about $2,000 below it. Further, this hump in the density function is especially pronounced among women who receive welfare income every month of a quarter—those for whom we would most expect it. The lack of precise bunching has been found elsewhere (e.g., see Emmanuel Saez's, 2002, study of tax rates).[22]

In sum, the evidence is consistent with behavioral responses to the Jobs First disregard policy. We cannot completely rule out the possibility that banking or queuing effects drive part of the observed negative earnings effect at the top of the earnings distribution. Evidence of such effects would be interesting and important in its own right, however, since both effects would suggest that time limits lead to lower-quality job matches.

B. *Exits from Administrative Data*

One concern in interpreting the QTE results above involves women who have zero total income in some quarters. For these women to survive, they must have some way to finance consumption other than UI-covered Connecticut earnings, cash assistance through Jobs First or AFDC, and food stamps. Such women could have some other source of earnings (UI-noncovered or under-the-table earnings); they could have support (cash or in-kind) from family members, absent noncustodial parents, or other government programs; or they could have moved out of Connecticut. A substantial amount of discussion in the final report, mostly using the three-year follow-up survey, suggests that neither marriage nor migration rates were systematically affected by welfare policies and that child support payments were only slightly impacted. That is not enough for our purposes, however, because it is always possible (for example) that high-earnings women systematically stayed in Connecticut because of Jobs First, while low-earnings women systematically moved out, a pattern that could affect QTE estimates. To deal with this issue, we consider the sample of women with zero total income in any quarter and find the last chronological quarter in which each had nonzero total income. We then exclude all subsequent quarters for such women from the analysis, which eliminates slightly more than a fifth of the sample of person-quarters. There is virtually no variation across treatment status in the overall probability of such attrition in the administrative data. Furthermore, at *each* quarter in the follow-up period, there are no statistically significant differences in the probability of exiting the sample between the treatment and control group. Nonetheless, we recalculated the QTE excluding our synthetic "movers." With the (expected) exception of parts of the distribution having zero income, the results estimated on this sample of nonmovers are qualitatively identical to the figures presented above.

C. *Issues Related to Child Support*

Jobs First changed both the amount of the paternal child support disregard and the way in which child support payments and benefit checks interact. Under AFDC rules, monthly child support checks for women receiving cash

[20] These predictions about features of the density around the poverty line assume that women understand the Jobs First disregard policy. Our discussions with Connecticut welfare officials indicate that the disregard policy was chosen in part because of its simplicity so that both recipients and caseworkers would understand the benefit formula.

[21] Data issues for this exercise include the fact that McCrary's (2005) results apply to continuous distributions while ours are rounded; earnings are quarterly rather than monthly; we do not have monthly data on family size; and even family size at intake is censored for families with more than three children.

[22] One application where evidence of bunching is more clear is in the Social Security earnings test, as shown by Steven J. Haider and David S. Loughran (2005) and, to a somewhat lesser extent, Leora Friedberg (2000).

assistance were to be paid to the state, which would then send the mother a child support check for up to \$50, with the state keeping the balance of the payment; the woman's welfare check was unaffected by the amount of the child support payment. Under Jobs First rules, the father's child support check is again sent to the state. However, the state now sends the mother a child support check equal to the exact amount the father sends the state. If the monthly child support payment is less than \$100, her welfare benefit check is unaffected; if the child support payment exceeds \$100, then the welfare benefit check is reduced by \$100 less than the amount of the child support payment. Thus Jobs First increases the monthly disregard by \$50 while changing the distribution of funds across the child support and welfare benefit checks. As we note in Section IB, we do not believe that the change in the disregard should have significant effects on labor supply for most women in the experimental sample.[23]

However, MDRC's public use file does not report monthly data on child support check amounts received by women. Since these checks can be much larger—with welfare benefit checks being correspondingly smaller except for the \$100 disregard—under Jobs First than under AFDC, we were concerned that our lack of child support data might cause our results on transfers and total income (as we are able to measure it) to be systematically biased. Whether this data asymmetry is practically problematic depends largely on the share of observations affected. We can assess how frequently women receiving Jobs First also received more than \$100 a month in child support, thanks to a quirk in the way MDRC constructed the public use file. MDRC first rounded (to the nearest \$50) the Jobs First benefit payment the woman would receive if her children's father were to pay \$100 or less in child support; call this amount B. Let the unrounded value of child support paid to the state by noncustodial fathers be S. The amount reported on MDRC's public use file equals $B - \max[0, S - 100]$, which is the difference between B and the unrounded value of nondisregarded child support paid to the state by noncustodial fathers.[24] It follows that observations on cash welfare payments to Jobs First women whose last two digits equal neither "50" nor "00" must have had child support exceeding \$100 during that month. Fewer than 8 percent of all quarterly Jobs First observations have unrounded values. Moreover, more than a third of the women with unrounded observations have either one or two quarters with unrounded values, and more than half have only four or fewer such quarters.

These facts suggest that our inability to measure transfers symmetrically is likely to be only a minor problem. To make sure, we conducted a number of sensitivity exercises, which confirmed that this data issue is unlikely to seriously affect our results. These exercises involved: (a) reestimating our QTE results for transfers and total income without the observations having unrounded values for the cash assistance variable; and (b) reestimating the results for transfers and total income using a number of imputation approaches to assigning child support values, based on data from MDRC's three-year followup survey. Because the details are lengthy, we do not describe these exercises in any detail here (a summary is available on request). A fair summary of the results of these exercises, however, is that they would lead to few substantive changes in our conclusions regarding QTE for transfers and measured total income. We note that our earnings results are necessarily unaffected since administrative data on earnings are completely unaffected by child support.

[23] Economists would not typically expect changes in "legal incidence"—how the total payment is broken down across the child support and welfare benefit—to affect labor supply. That said, the state's purpose in changing the distribution of payments across checks was to ensure that both women and noncustodial fathers could see how much child support was being paid on behalf of their children, in hopes that women's beliefs concerning the feasibility of leaving welfare would change, while men's willingness to send the state a check would increase. Daniel R. Meyer et al. (2003) report experimental evidence that Wisconsin's simultaneous implementation of both a full child support pass-through and a full disregard resulted in increased child support payments when compared to a partial pass-through and partial disregard.

[24] This algorithm was not explained in the public use file's documentation, but it was confirmed to us by an MDRC staff member in personal correspondence. Note also that this data quirk applies only to data for women assigned to the Jobs First group, since AFDC-group women's welfare checks are unaffected by the amount of child support paid on their behalf.

D. *Subgroups*

As noted in the introduction, the mean impacts literature has drawn the conclusion that there is little heterogeneity in treatment effects. However, some authors, e.g., Grogger and Karoly (2005, p. 231), have suggested that the common approach of using ad hoc subgroups would be unlikely to consistently reveal treatment effect heterogeneity even where it exists. To examine this issue, we followed a common approach in the welfare reform literature, considering separately high-school dropouts and women with at least a high-school diploma. High-school graduates are often used as a comparison group: given high-school graduates' lower welfare participation rates, reforms are often thought to affect them less than they do dropouts. To be part of the Jobs First experiment, all women in our sample at least had to apply for welfare, so this argument is less clear-cut than usual. Nonetheless, this is a logical way to consider the subgroups question.[25] We report detailed QTE results for dropouts and high-school graduates (including both those with and without postsecondary education) in our earlier working paper; here we simply summarize the main findings. First, differences in mean effects across dropout status are trivial. Second, heterogeneity in QTE within dropout status appears to be no less than the heterogeneity when we pool observations. Thus the most common approach based on mean impacts for judiciously chosen subgroups misses the entire heterogeneity story.

VII. Conclusion

Our results establish several clear conclusions. First, mean treatment effects miss a lot: estimated quantile treatment effects for earnings, transfers, and income show a great deal of heterogeneity. Theory predicts that mean treatment effects will average together opposing effects, and our results clearly confirm this prediction. Second, results for earnings are clearly consistent with predictions from labor supply theory that effects at the bottom should be zero, those in the middle should be positive, and (before time limits) those at the top should be negative. Third, the effects of Jobs First are very different in the pre– and post–time limit period, especially with respect to the transfers distribution. Negative effects at the top of the earnings distribution appear only in the pre–time limit period, as we would expect. This fact suggests a role for behavioral induced eligibility effects, most likely through reduced exit rather than increased entry. Banking and queuing effects are complementary explanations. Fourth, it is not unreasonable to believe that Jobs First led to substantial increases in income for a large group of women. On the other hand, once time limits take effect, Jobs First likely had at best no impact, and perhaps a negative one, on another sizable group of women. This finding is at odds with results in Schoeni and Blank (2003), who find positive effects throughout the distribution except in the very lowest percentiles. Moreover, we find that most of the positive shift in the income distribution occurs at above-median quantiles. Fifth, our results are robust to dropping observations from women who may have moved out of state or otherwise left the public assistance system while having no earnings (e.g., gotten married). Sixth, focusing on differences in mean treatment effects between dropouts and nondropouts—perhaps the most common comparison-group approach—is virtually useless in uncovering the treatment effect heterogeneity we demonstrate. In sum, our results show that QTE methodology can play a very useful role in assessing the effects of welfare reform when theory predicts heterogeneous treatment effects of opposing signs. We hope that this methodology will be used more often to address and analyze heterogeneous effects of welfare and other reforms.

Appendix A: Data Issues

Data are available for earnings (transfers) for eight (seven) quarters preceding random assignment. After random assignment, there are 16 quarterly observations on Connecticut earnings for every sample member except 30 people who entered the sample in January or February of 1997. Earnings data come from Connecticut's UI system, so earnings not covered by UI are missed; fortunately the vast majority of employment is

[25] Various parts of the final report (especially Appendix I) contain analyses of a wide array of subgroups. MDRC's focus is on groups labeled "most disadvantaged" and "least disadvantaged," which are defined using dropout status and employment and welfare use histories. We discuss these definitions in more detail in Bitler et al. (2003b).

covered by UI. Data on food stamps and welfare payments come from Connecticut's Eligibility Management System (EMS), which warehouses information about welfare use. To preserve confidentiality, MDRC rounded several key variables before releasing the public-use data (they rounded quarterly earnings data to the nearest \$100 and Jobs First, AFDC, and food stamps payments to the nearest \$50). For cases with true amounts between zero and the lowest reported nonzero value (\$50 or \$100), true values are rounded up, so that there are no false zeroes in the data.

Appendix B: Technical Issues

In this appendix, we present a brief discussion of some technical issues as well as descriptions of our various bootstrap procedures.

Results for inference on quantile treatment effects are well-known when the outcome variable is continuous (in which case sample quantiles—and thus QTE—are known to be asymptotically normal). In our case, there are two ultimately minor complications. First, our data are discrete due to MDRC's rounding. Second, to ensure finite-sample balance across all observable pretreatment variables, we weight each observation by its inverse propensity score. Gelbach (2005) argues that rounding is practically unimportant: sample quantiles are consistent for the population quantiles of the rounded random variable, and the empirical bootstrap distribution consistently estimates the sampling distribution of sample quantiles. Moreover, as Gelbach (2005) points out, Jeffrey M. Wooldridge's (2005) results on inverse propensity score weighting for *M*-estimators implies that such weighting does not affect these conclusions. Thus (a) our inverse propensity score weighted sample quantiles are consistent for the sample quantiles of each distribution; (b) our QTE are consistent estimates of the population QTE based on the rounded variables; and (c) the bootstrap percentile method that we use to estimate 90-percent confidence intervals consistently estimates the quantiles of the asymptotic sampling distribution for each individually estimated QTE. Thus, our empirical approach allows for consistent inference on the individual QTE.

Here is a very brief summary of the empirical method just discussed (we consider only the outcome variable earnings for quarters 1–7 in this discussion, but the same method is used for all three dependent variables and both time periods):

1. For each program group $t \in \{0, 1\}$, observed earnings values are sorted, lowest to highest.
2. For each person i and quarter s, we create the variable $\hat{F}_{i,s,t}(y) \equiv N^{-1} \sum_{i=1}^{n} \sum_{s=1}^{7} \hat{\omega}_i 1(T_i = t)1(Y_{i,s} \leq y)$, where n is the total number of people in the sample; and $N = 7n$ is the total number of person-quarters; $\hat{\omega}_i$ is the estimated inverse propensity score given person i's observable characteristics. To estimate the weights, we use predicted values from a logit model in which the treatment dummy is related to the following variables: quarterly earnings in each of the 8 preassignment quarters, separate variables representing quarterly AFDC and quarterly food stamps payments in each of the 7 preassignment quarters, dummies indicating whether each of these 22 variables is nonzero, and dummies indicating whether the woman was employed at all or on welfare at all in the year preceding random assignment or in the applicant sample. We also include dummies indicating each of the following baseline demographic characteristics: being white, black, or Hispanic; being never married or separated; having a high-school diploma/GED or more than a high-school education; having more than two children; being younger than 25 or age 25–34; and dummies indicating whether baseline information is missing (as is the case for fewer than 200 observations) for education, number of children, or marital status. Denoting the estimated propensity score for person i as $\hat{p}_i$ and the treatment dummy as T_i, the estimated inverse-propensity score weight for person i is

$$\text{(B1)} \qquad \hat{\omega}_i \equiv \frac{T_i}{\hat{p}_i} + \frac{1 - T_i}{1 - \hat{p}_i}.$$

3. Each element of our set of estimated sample quantiles $\{\hat{y}_{q,t}\}_{q=1}^{97}$ is defined as $\hat{y}_{q,t} \equiv \inf\{y : \hat{\mathrm{F}}_{\mathrm{i,s,t}}(y) \geq q\}$.
4. The estimated QTE for quantile q is then $\hat{\Delta}_q \equiv \hat{y}_{q,1} - \hat{y}_{q,0}$.
5. To estimate the sampling distribution of the set of estimated QTE, we use 1,000

APPENDIX TABLE 1—PERCENTILES OF THE DISTRIBUTION OF QUARTERLY EARNINGS, TRANSFERS, AND TOTAL INCOME FOR AFDC AND JOBS FIRST GROUPS

	Percentiles of the distribution								
	10	20	30	40	50	60	70	80	90
Earnings, Q1–7									
Jobs First group	0	0	0	0	100	800	1,500	2,500	3,700
AFDC group	0	0	0	0	0	300	1,200	2,200	3,900
Earnings, Q8–16									
Jobs First group	0	0	0	0	1,000	2,000	3,000	4,000	5,600
AFDC group	0	0	0	0	200	1,400	2,800	4,100	5,700
Transfers, Q1–7									
Jobs First group	0	0	1,050	1,650	1,800	1,950	2,176	2,400	2,732
AFDC group	0	0	400	1,050	1,550	1,800	1,950	2,250	2,700
Transfers, Q8–16									
Jobs First group	0	0	0	0	0	450	1,000	1,800	2,300
AFDC group	0	0	0	0	150	650	1,500	1,950	2,400
Total income, Q1–7									
Jobs First group	150	1,450	1,800	2,100	2,400	2,850	3,400	4,100	5,150
AFDC group	150	1,350	1,800	1,950	2,250	2,500	2,850	3,400	4,400
Total income, Q8–16									
Jobs First group	0	150	1,000	1,800	2,400	3,050	3,800	4,650	5,900
AFDC group	0	200	1,300	1,900	2,350	2,850	3,500	4,500	5,900

Notes: All percentiles calculated using inverse propensity-score weighting. Difference between Jobs First group percentile and AFDC group percentile is the QTE for the given percentile, graphed in Figures 3–8.

replications of the following bootstrap procedure:

(a) Randomly draw n persons with replacement from the sample, including data on all pretreatment observables and every quarter of earnings data for each selected person. Thus if person i is selected k times, we add k copies of i's earnings in each of the first seven quarters to the data (this resampling scheme is called the block bootstrap).

(b) Estimate the propensity score model on this sample of persons and compute the estimated weights $\hat{\omega}^*_{i,b}$ (where the asterisk denotes that this is a bootstrap statistic and b denotes that this is the b^{th} bootstrap iteration).

(c) Estimate the sample QTE $\hat{\Delta}^*_{q,b}$ for this iteration as described in the steps above, treating the person-quarter as the unit of observation.

6. Sort the 1,000 estimates of $\hat{\Delta}^*_{q,b}$, highest to lowest, with the b^{th} order statistic being $\Delta^*_{q(b)}$. Our 90-percent confidence interval is then $[\Delta^*_{q(950)}, \Delta^*_{q(51)}]$.

To test whether we can reject that all QTE in a range are either nonnegative or nonpositive (as in footnote 16 and Section VC), we use a method suggested by Abadie (2002) to impose the null that all QTE are exactly zero (we are grateful to a referee for suggesting this idea). Let n_0 be the number of real-data persons assigned to the control group, and let $n_1 = n - n_0$ be the number of real-data persons assigned to the treatment group. Here is the procedure we use to construct the null sampling distribution:

1. Randomly draw 1,000 size-n samples of persons from the data.
2. Assign a uniformly distributed random number to the i^{th} person in the b^{th} bootstrap sample, and sort the sample of persons by this random number. Assign $D^*_{i,b} = 0$ to the first n_0 persons in the b^{th} sample, and assign $D^*_{i,b} = 1$ to the remaining n_1 persons in this bootstrap sample.
3. Using the procedures described in steps 5(b), 5(c), and 6 above, estimate the q^{th} sample quantile for the b^{th} bootstrap sample of observations having $D^*_{i,b} = d$; call these sample quantiles $\tilde{y}^*_{q,b,d}$. The b^{th} null estimate $\tilde{\Delta}^*_{q,b}$ of Δ_q is then defined as $\tilde{\Delta}^*_{q,b} \equiv \tilde{y}^*_{q,b,1} - \tilde{y}^*_{q,b,0}$.

The tests mentioned in footnote 16 are as follows. First, we test whether the largest negative QTE in the set $\{\hat{\Delta}_q^{earnings}\}_{q=88}^{Q}$ is negative, where we consider each $Q \in \{96, 97\}$ (we consider both cases because the ninety-seventh quantile exhibits a good deal more variability in its tails than do the others). Using the real data, our largest-magnitude negative QTE in either range is $\hat{\Delta}_{92} = -300$; this is our test statistic's realized value. We are testing the null hypothesis that all QTE between quantiles 88–Q are nonnegative, so we consider a one-sided test. Of all the bootstrap QTE in the range 88–96, 7.0 percent have a negative value equal to 300 or greater in magnitude, while 11.0 percent in the range 88–97 do. Thus this test rejects at levels 0.07 or 0.11, depending on the range used. Our second test involving the pre–time limit earnings QTE is to compare the real-data number of negative QTE in the two ranges to the bootstrap null distribution for the number of negative QTE. The real-data value is 9 (10) when $Q = 96$ ($Q = 97$). Based on the bootstrap distribution, 14.1 percent (12.9 percent) of the bootstrap draws have this number of negative QTE. Thus we cannot reject at conventional levels [$p = 0.141(0.129)$].

We repeat these two basic testing procedures to test the nulls of no negative (no positive) QTE for quantiles 19–38 (48–88) of the total income distribution for quarters 8–16 (see Section VC); we discuss results in the text.

REFERENCES

Abadie, Alberto. "Bootstrap Tests of Distributional Treatment Effects in Instrumental Variable Models." *Journal of the American Statistical Association*, 2002, *97*(457), pp. 284–92.

Abadie, Alberto; Angrist, Joshua D. and Imbens, Guido. "Instrumental Variables Estimates of the Effect of Subsidized Training on the Quantiles of Trainee Earnings." *Econometrica*, 2002, *70*(1), pp. 91–117.

Adams-Ciardullo, Diana; Bloom, Dan; Hendra, Richard; Michalopoulos, Charles; Morris, Pamela; Scrivener, Susan and Walter, Johanna. *Jobs First: Final report on Connecticut's welfare reform initiative*. New York: Manpower Demonstration Research Corporation, 2002.

Ashenfelter, Orley. "Determining Participation in Income-Tested Social Programs." *Journal of the American Statistical Association*, 1983, *78*(383), pp. 517–25.

Bane, Mary Jo and Ellwood, David T. *Welfare realities: From rhetoric to reform*. Cambridge, MA: Harvard University Press, 1994.

Bitler, Marianne P.; Gelbach, Jonah B. and Hoynes, Hilary W. "Some Evidence on Race, Welfare Reform, and Household Income." *American Economic Review*, 2003a (*Papers and Proceedings*), *93*(2), pp. 293–98.

Bitler, Marianne P.; Gelbach, Jonah B. and Hoynes, Hilary W. "What Mean Impacts Miss: Distributional Effects of Welfare Reform Experiments." National Bureau of Economic Research, Inc., NBER Working Papers: No. 10121, 2003b.

Blank, Rebecca M. "Evaluating Welfare Reform in the United States." *Journal of Economic Literature*, 2002, *40*(4), pp. 1105–66.

Blank, Rebecca M.; Card, David E. and Robins, Philip K. "Financial Incentives for Increasing Work and Income among Low-Income Families," in David E. Card and Rebecca M. Blank, eds., *Finding jobs: Work and welfare reform*. New York: Russell Sage Foundation, 2000, pp. 373–419.

Blank, Rebecca M. and Schoeni, Robert F. "Changes in the Distribution of Children's Family Income over the 1990s." *American Economic Review*, 2003 (*Papers and Proceedings*), *93*(2), pp. 304–08.

Blinder, Alan S. and Rosen, Harvey S. "Notches." *American Economic Review*, 1985, *75*(4), pp. 736–47.

Bloom, Dan; Hendra, Richard; Kemple, James J.; Morris, Pamela; Scrivener, Susan and Verma, Nandita. *The family transition program: Final report on Florida's initial time-limited welfare program*. New York: Manpower Demonstration Research Corporation, 2000.

Bloom, Dan and Michalopoulos, Charles. "How Did Welfare and Work Policies Affect Employment and Income: A Synthesis of Research." New York: Manpower Demonstration Research Corporation, 2001.

Bloom, Dan; Scrivener, Susan; Michalopoulos, Charles; Morris, Pamela; Hendra, Richard; Adams-Ciardullo and Walter, Johanna. *Jobs First: Final report on Connecticut's welfare reform initiative*. New York: Manpower Demonstration Research Corporation, 2002.

Eissa, Nada and Liebman, Jeffrey B. "Labor Supply Response to the Earned Income Tax Credit." *Quarterly Journal of Economics*, 1996, *111*(2), pp. 605–37.

Firpo, Sergio. "Efficient Semiparametric Estimation of Quantile Treatment Effects." Unpublished Paper, 2005.

Fraker, Thomas; Moffitt, Robert and Wolf, Douglas. "Effective Tax Rates and Guarantees in the AFDC Program, 1967–1982." *Journal of Human Resources*, 1985, *20*(2), pp. 251–63.

Friedberg, Leora. "The Labor Supply Effects of the Social Security Earnings Test." *Review of Economics and Statistics*, 2000, *82*(1), pp. 48–63.

Friedlander, Daniel and Robins, Philip K. "The Distributional Impacts of Social Programs." *Evaluation Review*, 1997, *21*(5), pp. 531–53.

Gelbach, Jonah. "Inference for Sample Quantiles with Discrete Data." Unpublished Paper, 2005. Available at http://www.glue.umd.edu/~gelbach/papers/working-papers.html.

Grogger, Jeffrey T. "The Effects of Time Limits, the EITC, and Other Policy Changes on Welfare Use, Work, and Income among Female-Headed Families." *Review of Economics and Statistics*, 2003, *85*(2), pp. 394–408.

Grogger, Jeffrey and Michalopoulos, Charles. "Welfare Dynamics under Time Limits." *Journal of Political Economy*, 2003, *111*(3), pp. 530–54.

Grogger, Jeffrey T. and Karoly, Lynn A. *Welfare reform: Effects of a decade of change*. Cambridge, MA: Harvard University Press, 2005.

Haider, Steven J. and Loughran, David S. "Reconsidering Whether Labor Supply Responds to the Social Security Earnings Test." Unpublished Paper, 2005.

Heckman, James J., Smith, Jeffrey and Clements, Nancy. "Making the Most out of Programme Evaluations and Social Experiments: Accounting for Heterogeneity in Programme Impacts." *Review of Economic Studies*, 1997, *64*(4), pp. 487–535.

Imbens, Guido W. and Angrist, Joshua D. "Identification and Estimation of Local Average Treatment Effects." *Econometrica*, 1994, *62*(2), pp. 467–75.

McCrary, Justin. "Manipulation of the Running Variable in the Regression Discontinuity Design." Unpublished Paper, 2005.

McKinnish, Terra; Sanders, Seth and Smith, Jeffrey. "Estimates of Effective Guarantees and Tax Rates in the AFDC Program for the Post-OBRA Period." *Journal of Human Resources*, 1999, *34*(2), pp. 312–45.

Meyer, Daniel R.; Cancian, Maria; Caspar, Emma; Cook, Steven; Kaplan, Thomas and Mayer, Victoria. *W-2 child support demonstration evaluation phase 2: Final report*. Madison, WI: Institute for Research on Poverty, 2003.

Michalopoulos, Charles and Schwartz, Christine. *What works best for whom: Impacts of 20 welfare-to-work programs by subgroup*. New York: Manpower Demonstration Research Corporation, 2001.

Moffitt, Robert A. "Incentive Effects of the U.S. Welfare System: A Review." *Journal of Economic Literature*, 1992, *30*(1), pp. 1–61.

Moffitt, Robert A. "The Effect of Pre-PRWORA Waivers on Welfare Caseloads and Female Earnings, Income and Labor Force Behavior," in Sheldon H. Danziger, ed., *Economic conditions and welfare reform*. Kalamazoo: Upjohn Institute Press, 1999, pp. 91–118.

Moffitt, Robert A. "Welfare Programs and Labor Supply," in Alan J. Auerbach and Martin Feldstein, eds., *Handbook of public economics.* Vol. 4. Amsterdam: Elsevier Science, North-Holland, 2002, pp. 2393–2430.

Office of Family Assistance. *Temporary Assistance for Needy Families (TANF) fifth annual report to Congress*. Washington, DC: U.S. Department of Health and Human Services, 2003.

Saez, Emmanuel. "Do Taxpayers Bunch at Kink Points?" Unpublished Paper, 2002.

Schoeni, Robert F. and Blank, Rebecca M. "What Has Welfare Reform Accomplished? Impacts on Welfare Participation, Employment, Income, Poverty, and Family Structure." National Bureau of Economic Research, Inc., NBER Working Papers: No. 7627, 2000.

Wooldridge, Jeffrey M. "Inverse Probability Weighted Estimation for General Missing Data Problems." Unpublished Paper, 2005.