I Z A Institute of Labor Economics

Initiated by Deutsche Post Foundation

DISCUSSION PAPER SERIES

IZA DP No. 10905

Behavioral Responses and Welfare Reform: Evidence from a Randomized Experiment

Robert Paul Hartley
Carlos Lamarche

JULY 2017

I Z A Institute of Labor Economics

Initiated by Deutsche Post Foundation

DISCUSSION PAPER SERIES

IZA DP No. 10905

Behavioral Responses and Welfare Reform: Evidence from a Randomized Experiment

Robert Paul Hartley
Columbia University

Carlos Lamarche
University of Kentucky and IZA

JULY 2017

IZA – Institute of Labor Economics

Schaumburg-Lippe-Straße 5–9
53113 Bonn, Germany

Phone: +49-228-3894-0
Email: publications@iza.org

www.iza.org

ABSTRACT

Behavioral Responses and Welfare Reform: Evidence from a Randomized Experiment*

Recent studies have used a distributional analysis of welfare reform experiments suggesting that some individuals reduce hours in order to opt into welfare, an example of behavioral-induced participation. Using data on Connecticut's Jobs First experiment, we find no evidence of behavioral-induced participation at the highest conditional quantiles of earnings. We offer a simple explanation for this: women assigned to Jobs First incur welfare participation costs to labor supply at higher earnings where the control group is welfare ineligible. Moreover, as expected, behavioral components and costs of program participation do not seem to play a differential role at other conditional quantiles where both groups are eligible to participate. Our findings show that a welfare program imposes an estimated cost up to 10 percent of quarterly earnings, and these costs can be heterogeneous throughout the conditional earnings distribution. The evidence is obtained by employing a semi-parametric panel quantile estimator for a model that allows women to vary arbitrarily in preferences and costs of participating in welfare programs.

JEL Classification: J22, I38, C21, C33

Keywords: welfare reform, quantile regression, panel data, program participation

Corresponding author:
Carlos Lamarche
Department of Economics
University of Kentucky
223G Gatton College of Business & Economics
Lexington, KY 40506
USA
E-mail: clamarche@uky.edu

* We would like to thank Chris Bollinger, Matt Harding, Bill Hoyt, Jeff Smith, Eugenia Toma and Jim Ziliak for comments and insightful conversations. We are also thankful for comments and questions from workshop participants at the University of Kentucky. The data used in this paper were made available by MDRC. The authors remain solely responsible for how the data have been used or interpreted.

1. Introduction

Distributional effects of policies are increasingly the causal effect of interest among social scientists. For example, policy evaluation for experimental reforms may depend on considerations of target efficiency, such as whether program features induce behavioral responses where individuals reduce labor supply in order to become income-eligible for participation. Questions of equity and efficiency were prominent in the public debate leading up to the largest policy reforms in U.S. welfare history during the 1990s. By 1995, at least 40 states had implemented a policy waiver from the federal rules under Aid to Families with Dependent Children (AFDC) in order to experiment with work incentives and restrictions. Connecticut was a prime example for policy evaluation in that its reforms were a combination of the most generous earnings disregards (earnings not used in benefit reductions) and the most strict time limits. Beginning in 1996, Connecticut implemented a welfare waiver program called Jobs First, under which women could increase earnings up to the federal poverty line without any reduction in benefits. Welfare recipients and applicants were randomized into either Jobs First or AFDC, making the identification of quantile treatment effects (QTEs) possible by experimental design and under selection on observables. This leads to the key contribution of our study, which is to investigate empirically whether unobserved heterogeneity, possibly associated with welfare participation costs, is relevant for considerations of target efficiency.

The behavioral effects of welfare reform on low-income single mothers, especially regarding intensive margin responses on labor hours supplied, is an area of great interest in policy discussions and within the research literature. For the historical context and details of welfare policy reform from AFDC to the creation of Temporary Assistance for Needy Families (TANF), see Moffitt (2003) and Ziliak (2016). For a review of welfare reform effects across 15 European countries, see Immervoll, Kleven, Kreiner, and Saez (2007). Bargain (2006), Brewer, Duncan, Shephard, and Suárez (2006), and González (2008) investigate other welfare reform responses in Europe. Beffy et al. (2016) demonstrate the importance of restrictions on offered hours for low-income mothers in the United Kingdom. Mogstad and Pronzato (2012) compare work hours for married and single mothers after welfare reform in Norway. Similar work is beginning to look at behavioral responses to transfer programs in developing economies where the informal labor market provides an alternative margin of response (Bergolo and Cruces 2016 and Banerjee, Hanna, Kreindler, and Olken 2016).

It has been recognized in the literature that AFDC-assigned women leave welfare at different rates over time than those assigned to Jobs First since these programs have different features including earnings disregards and time limits (Bitler, Gelbach, and Hoynes 2006). It is also known that welfare imposes costs on participants including transactions costs and stigma (Moffitt, 1983, Currie 2006, Blank, Card and Robins 2000, Blank and Ruggles 1996). It is natural then to expect that these factors vary by treatment status after the random assignment, leading to a framework where AFDC-assigned women who exited welfare have no participation costs and Jobs First-assigned women who did not exit welfare have non-zero participation costs. In order to address the plausible cost differentials by treatment status, we use a semiparametric quantile estimator for a sparse econometric model and revisit the distributional analysis of the Jobs First welfare reform experiment.[1] Keys to the identification and estimation of QTEs are the experimental data we use and an assumption of zero participation costs for women who exited welfare at conditionally high earnings.

Consistent with the literature, we find a treatment effect estimate at the 0.90 quantile of -200 dollars of quarterly earnings which suggests that women reduced hours to become income-eligible for participation. In contrast, once we allow for costs of welfare participation to be different by individuals and treatment status (e.g., zero participation costs for AFDC-assigned women), we find evidence suggesting a positive treatment effect of 300 dollars. We propose and perform a test that indicates that these QTEs are significantly different at standard levels. At other quantiles, however, there is no statistically significant difference between estimates. To put the size of estimated treatment effect differences in context, Moffitt (1983) estimates that AFDC participation imposes a cost of about 4 hours of labor supply per week, which corresponds to around 520 dollars per quarter at an hourly wage of 10 dollars (the 90th percentile wage at quarter 1).[2] We interpret this evidence as indicative that women make choices regarding work and welfare based on their opportunity cost of time, which is increasing in earnings and depends on family structure, preferences, and program features given treatment status. Further, we expand the empirical analysis by estimating QTEs for continuing welfare recipients and new applicants given possible differences in participation costs. Long-term welfare participation, associated with ongoing recipients, may imply higher informational costs to labor supply due to limited labor market experience or the effects of persistent stigma. We find evidence that controlling for latent individual

[1] Our approach relates to recent penalized estimators for sparse models proposed in the literature (see, e.g., Belloni and Chernozhukov 2011, Imai and Ratkovic 2013, and Belloni, Chernozhukov, Fernández-Val and Hansen 2017; see also Harding and Lamarche 2017 for panel data). For theory and application of distributional analysis for welfare programs and treatment effects, see Heckman, Smith and Clements (1997), Bollinger, Gonzalez and Ziliak (2009) and Blundell, MaCurdy and Meghir (2007), among others. Burtless and Hausman (1978) and Moffitt (2002) investigate the importance of individual parameters in the specification of econometric models.

[2] Meyer and Rosenbaum (2001) incorporate variable stigma/transaction costs in a structural model with panel data and estimate that the monetized cost of working while on welfare is approximately \$643 per quarter.

heterogeneity affects behavioral-induced participation more for those with less labor market experience and longer, more frequent welfare spells compared to new applicants.

Heterogeneous impacts of the Jobs First experiment have already received notable attention. Bitler et al. (2006) illustrate the importance of estimating the distributional effects of a welfare reform experiment by using a nonparametric estimator for the difference between the treatment and control distributions at given quantiles. According to their estimates, the reform had no impact at the lower tail of the conditional distribution of earnings, it increased the conditional median of earnings, and it reduced the upper tail of the earnings distribution. While the mean treatment effect provides an uninformative summary of opposing effects, treatment effects exhibit significant differences across quantiles. More recently, Kline and Tartari (2016) estimate bounds for individual behavioral responses based on revealed preference assumptions. For women who would earn above the federal poverty line under AFDC, the authors estimate that at least 20 percent would reduce earnings in order to participate under Jobs First. Both of these studies find evidence that women in the upper earnings distribution reduce labor supply in order to receive welfare transfers, an example of behavioral-induced participation (Ashenfelter 1983). Consistent with these studies, we employ experimental data for Connecticut's Jobs First waiver program and estimate distributional effects. We distinguish our approach from the previous literature, however, by incorporating the panel nature of the Jobs First experiment and allowing a behavioral component of program participation. Our results compare directly to the nonparametric quantile treatment effect as constructed in Bitler et al. (2006), and we also discuss our design-based approach in comparison to the methodology and findings of Kline and Tartari (2016). The policy implications regarding behavioral-induced participation as described in the literature do not generalize to all individuals on welfare near the eligibility threshold, especially regarding continuing recipients.

The paper is organized as follows. Section 2 provides a simple framework for motivating the economic model and making testable predictions for heterogeneous labor supply responses to treatment. Section 3 introduces an econometric model consistent with the framework developed in Section 2 and then proposes a new approach for estimating QTEs in a regression setting. While Section 4 discusses the data, Section 5 presents the empirical analysis including regression results and inference on estimated quantile treatment effects. Section 6 offers a discussion on estimation issues and interpretation of results. Section 7 concludes.

2. Economic Implications of Jobs First and AFDC programs

Connecticut implemented Jobs First as a welfare reform waiver program beginning in 1996. Approximately half of the cash welfare participants were assigned to Jobs First and the other half to AFDC. The key feature of this waiver program is a 100-percent earnings disregard up to the federal poverty line, which leads to an implicit marginal tax of zero percent. In contrast, AFDC disregarded $120 of monthly earnings for the first year in the program and $90 after the first year. The statutory marginal tax rate on earnings under AFDC is 100 percent such that each additional dollar earned reduces transfers by one dollar. This dramatic policy change with respect to earnings creates a strong work incentive for many welfare participants, but it also creates a work disincentive around a significant notch at the federal poverty line above which Jobs First participants become income-ineligible for transfers.

While earnings disregards provide a strong motivation for predicting labor supply, the policy reform included other features that may be salient for behavioral responses. For instance, earnings were also fully disregarded in Food Stamp benefit determination for the treatment group. The experimental programs differed along other policy dimensions such us work requirements, sanctions, and time limits. While Jobs First has a strict 21-month time limit, AFDC has no time limits. Additionally, Jobs First participants were eligible for more generous transition benefits for child care and Medicaid after exiting welfare.[3] The program features of Jobs First demonstrate a range of policies implemented at the state level after the transition from AFDC to Temporary Assistance for Needy Families (TANF).

Bitler et al. (2006) note that static labor supply theory motivates the prediction of heterogeneous treatment effects across the distribution of total income (earnings plus cash welfare and Food Stamp benefits). Panel (a) in Figure 2.1 reproduces their stylized budget constraint with monthly income on the vertical axis and monthly work hours from 0 to M on the horizontal axis. The federal poverty line is indicated by FPL, and the welfare benefit guarantee amount by G. The Jobs First budget constraint with 100-percent earnings disregards is shown by segment AF, and the AFDC constraint is shown by segment AB.[4] Most women in the Jobs First study would begin at a location near point A, so the figure represents points where women might locate over time under the AFDC policy in order to predict how their behavior would change under Jobs First. In panel (a), women who would locate at points like D and E under AFDC might be induced to participate in welfare under

[3]See Bloom et al. (2002) for a detailed description of differences between Jobs First and AFDC programs.

[4]Whereas the AFDC statutory implicit tax rate is 100 percent, the effective tax rate would be somewhat lower in practice, as shown for segment AB.

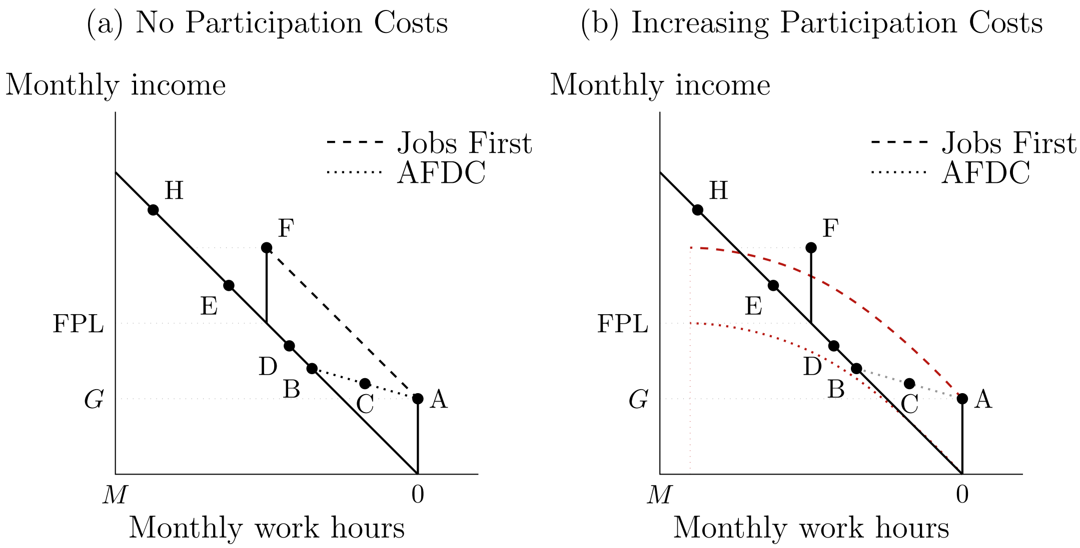

FIGURE 2.1. *Stylized Budget Constraints under Jobs First and AFDC. Panel (a) is reproduced from Bitler et al. (2006) Figure 1. Panel (b) shows a hypothetical constraint for Jobs First given nonlinear costs of welfare program participation.*

Jobs First. For instance, at point D, a woman under AFDC would become mechanically eligible for welfare under Jobs First and therefore might be induced to participate by income effect if leisure and consumption were normal goods. A woman located at point E under AFDC might be behaviorally induced to participate in Jobs First if the utility gain from participation compensates the reduced earnings necessary to become eligible below the federal poverty line.[5] If women in Jobs First are able to increase utility by reducing hours, this would imply a *negative* treatment effect at the upper tail of the earnings distribution.

The stylized budget constraint in Figure 2.1 illustrates points to which women may potentially relocate over time according to Bitler et al. (2006, pp. 994-995). Thus, treatment effect predictions presume that women in both experimental groups will likely increase earnings and relocate along their respective budget constraints. In order to observe behavioral-induced participation, the mechanism must either be reduced exits or reentry after exit. For both groups, there will be a natural rate of exit from welfare that may depend on program features. For instance, low-income women assigned to AFDC can locate at points like D or

[5]Ashenfelter (1983) emphasizes the distinction between behavioral components of participation, where an individual changes labor supply behavior in order to meet program eligibility, and mechanical components of participation, where changes in program rules make participation preferable given newly available choices in an individual's budget constraint.

E after random assignment, and consequently, one can expect a relative increase in participation due to Jobs First assignment. Therefore, policy impact estimates should take into account changes in work and participation decisions related to exposure to treatment over time as well as possible nonlinear changes in participation costs as hours increase. Becker (1965) motivates the relevance for one's cost of time with respect to labor market activities and household production. A woman's implicit time costs of working while also meeting welfare program requirements will depend on her characteristics such as family structure, neighborhood, preferences, or social capital.[6]

Suppose that women experience individual-specific costs of welfare participation that are increasing in hours of labor supply. Those with higher earnings have a higher opportunity cost of time, and thus transaction costs associated with welfare participation may be more relevant for women who are still eligible for welfare while working more hours.[7] Therefore, panel (b) suggests some hypothetical cost for participation that is increasing with labor hours such that a woman would have to work more hours to reach the same net earnings.[8] Note that women assigned to AFDC would not face such participation costs at higher earnings since participation eligibility is phased out at lower income levels. In this case, Jobs First would not be interpreted as a pure income shift at points like D and E, and thus the labor supply predictions become ambiguous depending on an individual's preferences and participation costs.

3. A Model and the Proposed Methodology

3.1. **Models.** To start analyzing the economic implications of the Jobs First program, we turn to the standard potential outcome approach to causal inference. A response variable or potential outcome has two values for a low-income woman i at quarter t, $(Y_{0,it}, Y_{1,it})$, one of which is observed and is labeled Y_{it}. The observed outcome depends upon the random treatment assignment, D_{it}, which can take $\{0, 1\}$ values indicating AFDC or Jobs First

[6]Given that the Jobs First and AFDC programs differ by other features besides earnings disregards and time limits, there might be other behavioral responses that could differ by treatment status over time.

[7]Also, Gottschalk (2005) shows that the preferences of low-income women may change with exposure to work and welfare participation, which implies that program features may influence women differently regarding the disutility of work or welfare.

[8]Moffitt (1983) illustrates stigma costs through lower levels of utility, though he notes that the utility model is "closely analogous to one in which costs of participation are monetized and included in the budget constraint." In Figure 2.1 panel (b), the dashed line indicates the implicit budget constraint (with monetized participation costs) for Jobs First. A dotted guideline is shown as parallel to net income by a height equal to the guarantee amount, G, and it continues up to an intersection with the FPL.

status, respectively. We then write $Y_{it} = D_{it}Y_{1,it} + (1 - D_{it})Y_{0,it}$ and assume that D_{it} is independent of the potential outcome.

Let $Q_Y(\tau)$ denote the τ-th quantile of the distribution of Y. The outcome variable Y is continuous and the treatment status D is independent of a p-dimensional vector of observed covariates, $\boldsymbol{x}$, as well as unobserved covariates, by the experimental design of the program. The parameter of interest is the quantile treatment effect (QTE) defined as

$$\Delta(\tau) = Q_Y(\tau|D = 1) - Q_Y(\tau|D = 0), \tag{3.1}$$

representing the change in earnings resulting from treatment at a given quantile. It is known that the QTE can also be obtained as a parameter of interest in a quantile regression model. It is immediately apparent that Y_{it} can be written as $Y_{it} = Y_{0,it} + \Delta D_{it}$ where $\Delta = Y_{1,it} - Y_{0,it}$. Therefore, $\Delta(\tau)$ can be obtained from the quantile model associated with the following equation: $Y_{it} = \Delta D_{it} + u_{it}$, where the error term is $u_{it} = Y_{0,it}$.[9]

As an extension to the previous model, consider now that the treatment is subject-specific satisfying $\Delta_i + Y_{0,it} = Y_{1,it}$. This can be motivated by individual costs associated with welfare participation. Without loss of generality, let $\Delta_i = v_i + \Delta$. It follows then that the treatment effect can be estimated using $Y_{it} = Y_{0,it} + v_i D_{it} + \Delta D_{it}$, or, if the treatment indicator is considered time-invariant given assignment at $t = 0$, then $Y_{it} = Y_{0,it} + \alpha_i + \Delta D_{i0}$, where unobserved individual heterogeneity is represented by $\alpha_i = v_i D_{i0}$. A distinctive feature of this extension is that it leads to a *sparse* model where α_i is equal to zero if $D_{i0} = 0$ and $\alpha_i = v_i$ if $D_{i0} = 1$, which leads to the following QTE parameter:

$$\tilde{\Delta}(\tau) := \Delta(\tau) + \alpha(\tau) = Q_Y(\tau|D = 1, \alpha) - Q_Y(\tau|D = 0). \tag{3.2}$$

The QTE parameter in equation (3.2), $\tilde{\Delta}(\tau)$, is identical to that in (3.1), $\Delta(\tau)$, in two cases. First, participation costs are not different by treatment status, implying that $Q_Y(\tau|D = 1, \alpha) - Q_Y(\tau|D = 0, \alpha) = \Delta(\tau)$. Note however that this case is not consistent with the empirical observation that women assigned to AFDC leave welfare at different rates over time than those assigned to Jobs First. Moreover, $\tilde{\Delta}(\tau) = \Delta(\tau)$ if Y_1 is independent of v. However, the assumption $Q_Y(\tau|D = 1, \alpha) = Q_Y(\tau|D = 1, v) = Q_Y(\tau|D = 1)$ contradicts the mechanism discussed in Figure 2.1 where participation costs v_i are not independent of $Y_{1,it}$.

At the upper quantiles of earnings, the model has a natural interpretation. Women assigned to AFDC would not face participation costs at higher earnings (since they are not

[9]Naturally, the model includes an intercept, but it is omitted here to simplify the presentation of the model and parameter of interest.

likely to participate), while women assigned to Jobs First do face participation costs. Returning to Figure 2.1, a woman i that would locate, for example, at point E (above the poverty line) would only reduce hours if the decreased earnings plus her individual cost of welfare participation, $\alpha_i = v_i$, are compensated by the increased transfers under Jobs First. Therefore, there is a range of earnings toward the upper conditional quantiles where welfare participation costs are relevant to evaluate behavioral-induced effects.

To evaluate the role of participation costs, we briefly focus on two particular quantile functions and discuss their interpretation within the literature.

Example 1. Consider the following simple model for earnings similar to those estimated in the literature: $Y_{it} = \beta + \Delta D_{i0} + \alpha_i + (1 + D_{i0}\gamma_0 + \alpha_i\gamma_i)u_{it}$, where γ_0 is a scale parameter, γ_i is a function of the transfer benefit, and the error term is distributed as F with location zero and unit variance. The corresponding quantile function is $Q_{Y_{it}}(\tau|D_{i0}, \alpha_i) = \beta(\tau) + \Delta(\tau)D_{i0} + \alpha_i(\tau)$, where $\beta(\tau) = \beta + F_u^{-1}(\tau)$, $\Delta(\tau) = \Delta + \gamma_0 F_u^{-1}(\tau)$ and $\alpha_i(\tau) = \alpha_i\left(1 + \gamma_i F_u^{-1}(\tau)\right) = v_i D_{i0}\left(1 + \gamma_i F_u^{-1}(\tau)\right)$. If $D_{i0} = 0$, then $\alpha_i(\tau) = 0$, representing the case of no participation costs. If $D_{i0} = 1$, then $\alpha_i(\tau) = v_i(1 + \gamma_i(\tau))$, which is expected to be increasing on τ. In this model, v_i might be interpreted as 'flat' stigma and $\gamma_i(\tau)$ as 'variable' stigma (Moffitt 1983).[10]

Example 2. Consider the model presented in Example 1 where participation affects hours offered, and consequently it affects earnings. For simplicity, assume that scale parameters (γ_0, γ_i) are zero for all mothers. We then have $Y_{it} = \beta + \Delta D_{i0} + \alpha_i R_i + u_{it}$, where R_i is an indicator variable for whether woman i participates on welfare. Note that the associated conditional quantile function includes an additional variable R_i and omission of this variable in a quantile model leads to inconsistent results since participation at t might not be independent of D_{i0}. Joint decisions for welfare participation and labor supply are discussed extensively in the literature (see, e.g., Moffitt 1983, 2002, among others).

This analysis leads to simple tests that can be performed using regression analysis as follows. We first estimate the QTE in equation (3.1) and then compare with the QTE in

[10] Moffitt (1983) proposes an economic model where 'flat' stigma is defined as the cost related to any welfare participation, and 'variable' stigma is a cost proportional to the size of the benefit. Although the benefit size is constant for Jobs First participants, variable stigma may be considered the cost proportional to the benefit-to-income ratio, which is decreasing over the earnings distribution. While Moffitt found no evidence of variable stigma, he provided mean estimates that were pre-welfare reform under AFDC such that any variable costs of participation at higher incomes post-welfare reform are unknown, such as the case of Jobs First.

equation (3.2). The difference between parameters is:

$$C(\tau) := \tilde{\Delta}(\tau) - \Delta(\tau) \geq 0, \tag{3.3}$$

and it represents the relative cost of welfare participation at different quantiles of earnings. At low conditional quantiles of earnings, we expect $C(\tau) = 0$ because α is zero or the cost of participation does not vary by treatment status, and at high conditional quantiles, we expect $C(\tau) > 0$ because Jobs First-assigned women incur participation costs to labor supply but AFDC-assigned women are not eligible for welfare, and consequently have zero cost of participation. Additionally, we estimate the QTE in equation (3.1) and then compare with the QTE of a model that incorporates an indicator variable for participation. In the next section, we discuss estimation approaches for the QTE parameter and we turn to investigating and testing the previous hypotheses in Sections 5.3 and 5.5.

3.2. **Background.** Although the QTE in equation (3.1) can be estimated using different approaches (see, e.g., Koenker 2005), previous analyses of welfare reforms have been concerned with potential selection into treatment. For consistent estimation of the parameter of interest, $\Delta(\tau)$, the identification restriction used in the literature is known as selection on observables (Rubin 1977, Heckman, Ichimura, Smith and Todd 1998, Firpo 2007). Thus, given a set of covariates, it is typically assumed that women are randomly assigned to Jobs First or AFDC. This identification condition gives rise to different estimation strategies.

A nonparametric approach, denoted here as NP, to estimating the QTE for individuals $i = 1, \ldots, N$ pooled over quarters $t = 1, \ldots, T$ is given by

$$\hat{\Delta}(\tau) = \inf\left\{y : \hat{F}_{Y_{it}}(y|D_{i0} = 1) \geq \tau\right\} - \inf\left\{y : \hat{F}_{Y_{it}}(y|D_{i0} = 0) \geq \tau\right\},$$
$$\text{for } \hat{F}_{Y_{it}}(y|D_{i0} = d) = \left[\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\hat{w}_i(\boldsymbol{x}_i) \cdot 1(D_{i0} = d) \cdot 1(y_{it} \leq y)\right], \tag{3.4}$$

where $d = \{0, 1\}$ and the empirical inverse-propensity weight is $\hat{w}_i(\boldsymbol{x}_i) = D_{i0}/\hat{p}_i(\boldsymbol{x}_i) + (1 - D_{i0})/(1 - \hat{p}_i(\boldsymbol{x}_i))$. This approach was used by Bitler et al. (2006). The variable $\hat{p}_i(\boldsymbol{x}_i)$ is the estimated propensity score obtained from a logit regression of an individual's propensity to be treated conditional on observed characteristics $\boldsymbol{x}_i$. Also considering selection on observables, Firpo (2007) proposes an estimation method that is a weighted version of the classical quantile regression estimator for cross-sectional data (Koenker and Bassett 1978).[11] The method is semiparametric in the sense that no parametric assumption is made on the

[11]Alternative program evaluation methodologies have been recently proposed by Cattaneo (2010) and Słoczyński and Wooldridge (2016). Our treatment, however, is not multivalued and we employ inverse-propensity score weighting as in Bitler et al. (2006) for comparability of results.

joint distribution of the observed variables. The quantile regression estimator (QR) for the QTE in equation (3.1) can be obtained by solving:

$$\min_{(\beta_0,\Delta)\in\boldsymbol{\Theta}} \sum_{i=1}^{N}\sum_{t=1}^{T} \hat{w}_i\left(\boldsymbol{x}_i\right)\cdot\rho_\tau\left(Y_{it}-\beta_0-\Delta D_{i0}\right), \tag{3.5}$$

where $\rho_\tau(u) = u\left(\tau - I\left(u \leq 0\right)\right)$ is the standard piecewise-linear quantile check function. We consider two variations of the minimizer of equation (3.5): a weighted estimator defined as above and an unweighted estimator with $\hat{w}_i(\boldsymbol{x}_i) = 1$ for AFDC and Jobs First participants.

3.3. **A Penalized Semiparametric Approach.** Identification of the time-invariant treatment effect is based on the sparse nature of the model. Let N_0 denote the number of women assigned to the control group, and N_1 the number assigned to treatment, with $N_0 + N_1 = N$. Thus, the model can be augmented by $\boldsymbol{\alpha} = (\boldsymbol{\alpha}_0', \boldsymbol{\alpha}_1')'$, where $\boldsymbol{\alpha}_0$ is an N_0-dimensional sparse vector for the set of individuals in AFDC, $i \in \mathcal{D}_0 = \{i : D_{i0} = 0\}$, and $\boldsymbol{\alpha}_1$ is an N_1-dimensional vector of individual effects for the set of individuals in Jobs First, $i \in \mathcal{D}_1 = \{i : D_{i0} = 1\}$. This model can be estimated by an extension of existing penalized quantile estimators with individual effects since the conditions for consistent estimation are satisfied by the experimental design of the program (Assumption 2, Lamarche 2010). We augment the QR optimization problem defined in equation (3.5) with individual effects and we introduce an ℓ_1 penalty function of the following form:

$$Pen(\alpha) = \sum_{i=1}^{N}|\alpha_i| = \sum_{i=1}^{N}(D_{i0}|\alpha_{i,1}| + (1-D_{i0})|\alpha_{i,0}|) = \sum_{j\in\mathcal{D}_0}|\alpha_j| + \sum_{k\in\mathcal{D}_1}|\alpha_k|. \tag{3.6}$$

Recall that the framework discussed before leads to a model with $N_1 < N$ individual effects possibly having non-zero costs of welfare participation. Following the economic intuition in Sections 2 and 3.1, individual effects enter into the equation via a sparse relationship to treatment, that is $\alpha_i = v_i$ for women in Jobs First and $\alpha_i = 0$ for those in AFDC. In order to incorporate this restriction in a penalized setting, we allow the Tikhonov regularization parameter, or tuning parameter, to be defined by treatment status, $\{\lambda_0, \lambda_1\}$. The regularization parameter shrinks the influence of individual effects toward zero as λ_d increases, so imposing the condition $\lambda_0 \gg \lambda_1$ would imply that participation costs are smaller in the control group than in the treatment group.[12] (The selection of the tuning parameters for the methods presented in this section is discussed in the Technical Appendix.) Therefore,

[12]Imai and Ratkovic (2013) use separate constraints in a least absolute shrinkage and selection operator (LASSO) framework in order to select pre-randomization variables that are causally related to heterogeneous treatment responses, though their study does not address quantiles nor the panel dimension.

the restricted panel quantile regression estimator (R-PQR) at a given τ can be estimated by solving

$$\min_{(\beta_0,\Delta,\boldsymbol{\alpha}')\in\boldsymbol{\Gamma}} \sum_{i=1}^{N}\sum_{t=1}^{T} \hat{w}_i(\boldsymbol{x}_i)\cdot\rho_\tau\left(Y_{it}-\beta_0-\Delta D_{i0}-\alpha_i\right)+\lambda_0\sum_{j\in\mathcal{D}_0}|\alpha_j|+\lambda_1\sum_{k\in\mathcal{D}_1}|\alpha_k|\,. \tag{3.7}$$

The restricted estimator can be easily adapted to the simultaneous estimation of J conditional quantiles as in Koenker (2004), and it differs from existing penalized estimators for panel quantiles (e.g., Harding and Lamarche 2017) by controlling for selection on observables. Moreover, the proposed semiparametric estimator in equation (3.6) relaxes the identification conditions in the literature by partially allowing for selection on time-variant unobservables when $\lambda_0=\lambda_1=c$.

More importantly, the restricted estimator offers a direct test of the economic model. It is designed to allow for program-specific costs of participation that are more relevant to Jobs First-assigned women than to AFDC-assigned women at conditionally higher earnings.

In theory, AFDC-assigned women at the upper quantiles of monthly income do not have costs of participating since they are out of welfare, while Jobs First-assigned women may have costs of participating since they are still welfare-eligible. Although in practice AFDC-assigned women were in general out of welfare several quarters after the reform, some of them were still in the program. Moreover, stigma effects can persist over time. Therefore, it is important to consider the case of $\lambda=\lambda_0=\lambda_1$, which does not strictly impose the framework presented in Section 2, however it does shrink the smallest individual effects, presumably the $\hat{\alpha}_{i_0}$'s, to zero.

Let the unrestricted panel quantile regression (PQR) estimator of the QTE at a given τ be defined as

$$\min_{(\beta_0,\Delta,\boldsymbol{\alpha}')\in\boldsymbol{\Gamma}} \sum_{i=1}^{N}\sum_{t=1}^{T} \hat{w}_i(\boldsymbol{x}_i)\cdot\rho_\tau\left(Y_{it}-\beta_0-\Delta D_{i0}-\alpha_i\right)+\lambda\sum_{i=1}^{N}|\alpha_i|\,, \tag{3.8}$$

where $\lambda\in\mathbb{R}_+$ is a tuning parameter. Notice that for identification of the QTE, the tuning parameter cannot be equal to zero. But for small values of λ, the QTE should be interpreted as an estimator from a model with individual effects. As the value of λ increases, the individual effects go to zero such that PQR estimates converge to QR estimates. If the restricted and unrestricted estimators give similar results, then the evidence supports a differential role of individual effects by treatment status, which corresponds to the suggested extensions to the economic framework in Section 2 and modeling assumptions above.

4. Data

Experimental data for Connecticut's Jobs First waiver program were obtained from MDRC (formerly Manpower Demonstration Research Corporation). The observations represent 4803 women, current welfare recipients or new applicants, who were randomly assigned into either the AFDC control group ($N_0 = 2407$) or the Jobs First treatment group ($N_1 = 2396$). Along with a range of time-invariant demographics, the data include quarterly measures of earnings rounded to the nearest hundred dollars, and AFDC and Food Stamps benefits rounded to the nearest fifty. Income measures are observed for quarters -8 to -1 before random assignment, quarter 0 at the time of random assignment into treatment (D_{i0}), and quarters 1 to 16 where the Jobs First time limit binds by quarter 7.[13]

Although Jobs First was a randomized experiment, a simple inspection of basic statistics reveals the presence of statistically significant differences by treatment status (Table 4.1). For instance, women in the Jobs First group have on average less quarterly earnings and more cash welfare. This however has been well documented in the literature. To address sample selection, Bitler et al. construct inverse-propensity weights by estimating the probability of treatment conditional on 60 variables including quarterly pre-treatment earnings and transfers as well as indicators for individual characteristics and family structure at the time of random assignment.[14] After estimating each individual's propensity to be treated, $\hat{p}_i$, the inverse-propensity weight is constructed as $\hat{w}_i = D_{i0}/\hat{p}_i + (1 - D_{i0})/(1 - \hat{p}_i)$ where as before D_{i0} indicates treatment status. For consistent comparison of results, we employ the same weights in the estimators defined in equations (3.4), (3.5), (3.7) and (3.8).

The descriptive statistics in Table 4.1 highlight the role of sample correction as well as some general outcomes of the treatment. First, note that randomization works well overall based on the unadjusted differences shown in the third column. The differences that remain statistically significant, though, may be important for estimating treatment effects on earnings. In addition to Jobs First participants having less earnings and more cash transfers

[13]Earnings data are missing for 30 women in quarter 16; AFDC and Food Stamps data are missing for 3175 women in pre-treatment quarter -8. We refer the reader to Bloom et al. (2002) for a background of the Jobs First program including, for example, experimental design, program implementation, and initial outcomes.

[14]The covariates are quarterly levels of pre-treatment earnings, cash transfers, and Food Stamps; quarterly indicators for any pre-treatment earnings, cash transfers, and Food Stamps; indicators for new applicant status at randomization, any employment in the year before randomization, any cash transfers in the year before randomization; indicators for black, Hispanic, white, never married, married/living apart, age less than 25, age 25-34, no high school degree or GED, high school degree or GED, more than two children; and, indicators for any missing data for education, children, and marital status.

Variables	Levels: Jobs First	Levels: AFDC	Differences: Unadjusted	Differences: Adjusted
Newhaven County (urban)	0.753	0.757	-0.004	-0.000
	(0.431)	(0.429)	(0.012)	(0.012)
Never married	0.624	0.631	-0.007	-0.000
	(0.484)	(0.483)	(0.014)	(0.016)
HS dropout	0.331	0.313	0.018	-0.000
	(0.471)	(0.464)	(0.013)	(0.013)
More than two children	0.227	0.206	0.021*	-0.000
	(0.419)	(0.405)	(0.012)	(0.012)
Mother younger than 25	0.289	0.297	-0.007	-0.000
	(0.454)	(0.457)	(0.013)	(0.014)
Mother older than 34	0.301	0.286	0.015	0.000
	(0.459)	(0.452)	(0.013)	(0.014)
Recipient (stock) sample	0.624	0.593	0.031*	-0.001
	(0.484)	(0.491)	(0.014)	(0.014)
Currently working $\geq$ 30 hours	0.276	0.313	-0.037	-0.033
	(0.447)	(0.464)	(0.029)	(0.027)
Hourly wage	6.583	6.808	-0.225	-0.164
	(2.234)	(2.592)	(0.155)	(0.153)
Public or subsidized housing	0.356	0.346	0.010	0.003
	(0.479)	(0.476)	(0.014)	(0.014)
Ever on AFDC as a child	0.248	0.258	-0.010	-0.010
	(0.432)	(0.438)	(0.013)	(0.013)
Ever received AFDC at prior quarter 7	0.548	0.528	0.020	-0.000
	(0.498)	(0.499)	(0.014)	(0.014)
Length in months of 1st AFDC spell	17.622	14.221	3.402*	3.115*
	(9.910)	(10.654)	(0.344)	(0.380)
Number of AFDC spells	1.173	1.217	-0.044*	-0.046*
	(0.583)	(0.685)	(0.018)	(0.018)
Long-term recipient ($>$ 2 years)	0.569	0.554	0.015	-0.005
	(0.495)	(0.497)	(0.014)	(0.013)
Pre-Treatment Quarters				
Average quarterly earnings	678.908	785.895	-106.988*	-0.887
	(1303.749)	(1544.720)	(41.240)	(108.313)
Average quarterly cash welfare	890.818	835.112	55.706*	-0.833
	(806.032)	(784.845)	(22.958)	(23.029)
Fraction of quarters with earnings	0.322	0.351	-0.029*	0.000
	(0.363)	(0.372)	(0.011)	(0.011)
Fraction of quarters with cash welfare	0.573	0.544	0.029*	-0.001
	(0.452)	(0.450)	(0.013)	(0.013)
Experimental Quarters 1-7				
Average quarterly earnings	1173.187	1139.047	34.141	81.931
	(1501.393)	(1739.033)	(46.875)	(123.695)
Average quarterly cash welfare	1083.255	889.050	194.205*	167.264*
	(620.003)	(639.856)	(18.180)	(20.162)
Fraction of quarters with earnings	0.514	0.450	0.064*	0.077*
	(0.394)	(0.398)	(0.011)	(0.012)
Fraction of quarters with cash welfare	0.746	0.662	0.084*	0.071*
	(0.345)	(0.380)	(0.010)	(0.011)

TABLE 4.1. *Descriptive Statistics by Treatment Status. Standard deviations are shown in parentheses, and * denotes statistically significant differences at the 10-percent level.*

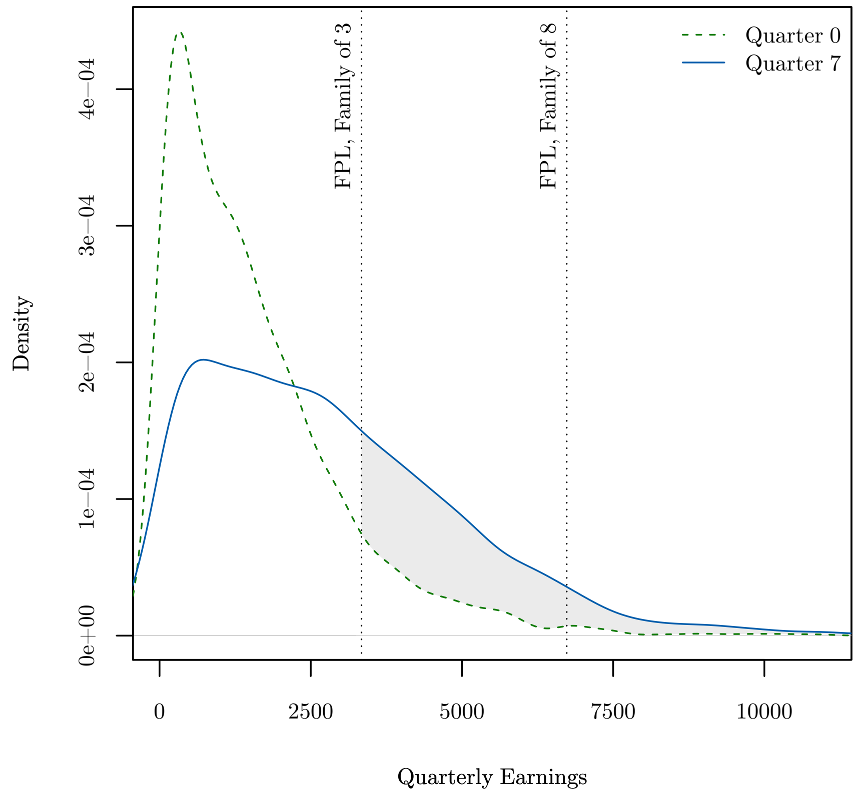

FIGURE 4.1. *Empirical Distributions of Earnings at Quarters 0 and 7. FPL denotes the federal poverty line as of 1997, the midpoint between random assignment and quarter 7.*

on average during pre-treatment quarters, they also tend to have larger families and are more likely to be continuing recipients instead of new applicants. Another distinction between Jobs First and AFDC participants is that the Jobs First participants have longer first spells on welfare yet slightly fewer spells, on average. The fourth column in Table 4.1 shows sample differences adjusted by inverse-propensity weighting. For the variables adequately controlled for in the first-stage propensity estimation, sample-corrected characteristics are well balanced: only very small and statistically insignificant differences remain in the adjusted variables for the pre-treatment period. However, for the exceptions of number and duration of previous spells, there are still statistically significant differences at the 10-percent level.

Of the 4803 total participants, 2923 women are continuing welfare recipients compared to 1880 who are new applicants. When the Jobs First experiment began in 1996, the entire state of Connecticut transitioned to Jobs First except for the two experimental counties, Newhaven and Manchester. Therefore, any new applicants would have different selection into welfare than continuing recipients. The same descriptive statistics discussed above are shown by participant type for recipients and applicants in Table A.1. As expected, applicants are different from recipients by every dimension shown in the table. In summary, applicants are less urban, more educated, less reliant on transfers, and they work more hours at higher wages. In particular, only about 20 percent of applicants had ever received AFDC around 2 years prior to random assignment compared to just over 75 percent of recipients, and the probability of being a long-term welfare recipient is about 43 percentage points lower for the applicant group on average. Applicants also earn more and receive less transfers than recipients both before and after random assignment. While applicants and recipients may experience some similar costs of welfare participation in terms of hassle, the ongoing recipients group may face additional costs related to persistent labor force detachment.

Lastly, Figure 4.1 shows the empirical distribution of earnings for working women at random assignment (quarter 0) and at the Jobs First time limit (quarter 7). The figure also shows the federal poverty line (FPL), which varies by family size. In the Jobs First data, the modal family size is approximately 3 members with the maximum size around 8.[15] The figure highlights how the probability that families earn more than the federal poverty line changes over time, which is shown by the area shaded in gray. Referring to Figure 2.1, the likelihood of locating around points E or H (locations at or above the federal poverty line) is small at the time of random assignment, but some participants will relocate to those higher earnings locations by quarter 7. The increase in individuals relocating near the poverty line, as shown in Figure 4.1, motivates the presumption that treatment effects at upper quantiles are related to potential behavioral responses around the eligibility notch.

5. Empirical Analysis

In this section, we employ the proposed estimation methods to investigate whether there is evidence that suggests that individuals reduce hours in order to opt into welfare, the behavioral-induced participation hypothesis. We compare our findings with results obtained by employing alternative estimation methods. Finally, we formulate a series of tests to examine whether the new results offered in this study are significantly different than existing results. Standard errors for all empirical results are constructed based on a block bootstrap method for comparability across estimators; for details, see Technical Appendix Section B.2.

[15]Administrative data on family size is available for only 225 individuals in both quarter 0 and quarter 7. Otherwise, family structure is identified by the variable *kidcount*, which is top-coded at 3.

5.1. **Pooled Data Results.** As a baseline estimate of the QTE, we present results based on the non-parametric approach given by equation (3.4) and the semiparametric approach introduced in equation (3.5). We restrict our attention to earnings in the first 7 quarters after the reform is introduced in order to focus on behavioral responses in the upper tail of the earnings distribution before the Jobs First time limit becomes binding. If behavioral-induced participation is expected, it would be most evident before time limits apply to women assigned to Jobs First. Estimates are also provided for total income (earnings, cash welfare, and Food Stamps) for quarters 8-16, which represent a long-run outcome.

Table 5.1 presents results for the QTE parameter given by the non-parametric estimator where estimates obtained with inverse-propensity weighting are shown in column (1) and estimates without weights in column (2). We reproduce all of the NP estimates exactly with only slight variations in the confidence intervals based on different random samples used for the 1000 bootstrap replications.[16] Given the random design of Jobs First, which remains a model program for welfare reform evaluation, one might expect the unweighted and weighted QTE results to be similar. In fact, there is no qualitative difference and only small quantitative differences in the point estimates presented in columns (1) and (2). For the results shown, the only difference from weighting the NP estimates is at the 0.75 quantile for total income in quarters 8-16: 300 in column (1) and 250 in column (2), though this difference is not statistically significant at conventional levels.[17]

Consistent with the predictions of the framework described in Figure 2.1 (panel (a)), the table shows that the reform had no impact at the lower tail of the conditional earnings distribution and it increased earnings at the 0.50 and 0.75 quantiles. At the upper tail, NP estimates indicate that the reform reduces earnings by 200 dollars, suggesting that some women reduced hours in order to opt for welfare.

Under the assumption that the treatment effect is linear and treatment status is randomly assigned, the nonparametric estimator for the QTE and the quantile regression estimator for a model that conditions on the treatment indicator variable are expected to yield similar results (Koenker 2005). The weighted and unweighted QR results are shown in columns

[16] In order to reproduce the confidence intervals reported in Bitler et al. (2006), it is necessary to use only bootstrap samples that are sufficiently balanced given that their software, which is available through the AER website, weights the empirical cumulative distributions of each treatment group across the full sample. This procedure causes the inference on the nonparametric approach to appear artificially more precise than otherwise with respect to the semiparametric quantile regression methods.

[17] Comparing weighted and unweighted QTEs for several quantiles between 0.05 and 0.95 (in results not shown here but available upon request), we find that there is no qualitative difference by weighting, and little quantitative difference.

τ	NP (1)	NP (2)	QR (3)	QR (4)	R-PQR (5)	R-PQR (6)	PQR (7)	PQR (8)
	Earnings, Quarters 1-7							
0.10	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
0.25	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
0.50	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00
	(31.92)	(30.92)	(32.88)	(28.30)	(52.64)	(52.64)	(52.66)	(52.66)
0.75	300.00	300.00	300.00	100.00	500.00	600.00	400.00	500.00
	(94.59)	(129.66)	(93.61)	(100.58)	(122.58)	(108.90)	(123.67)	(115.44)
0.90	-200.00	-200.00	-200.00	-300.00	300.00	400.00	300.00	200.00
	(117.68)	(217.92)	(119.89)	(128.17)	(110.08)	(120.12)	(110.49)	(107.01)
	Total Income, Quarters 8-16							
0.10	0.00	0.00	0.00	0.00	-600.00	-650.00	-350.00	-400.00
	(0.00)	(0.00)	(0.00)	(0.00)	(91.55)	(89.92)	(101.78)	(89.66)
0.25	-150.00	-150.00	-150.00	-150.00	-50.00	0.00	-50.00	-150.00
	(109.66)	(123.39)	(108.96)	(118.75)	(60.79)	(74.86)	(60.57)	(61.89)
0.50	50.00	50.00	50.00	50.00	50.00	50.00	50.00	50.00
	(64.62)	(74.46)	(64.08)	(67.64)	(75.60)	(75.60)	(75.61)	(75.62)
0.75	300.00	250.00	300.00	200.00	227.45	129.16	228.00	250.00
	(90.67)	(125.68)	(90.68)	(100.02)	(81.92)	(75.26)	(80.88)	(81.20)
0.90	0.00	0.00	0.00	0.00	474.17	573.99	300.00	328.00
	(118.30)	(213.37)	(118.33)	(120.37)	(99.22)	(99.06)	(87.10)	(87.35)
IPW	Yes	No	Yes	No	Yes	No	Yes	No

TABLE 5.1. *Quantile Treatment Effects on the Distributions of Earnings and Total Income. NP denotes the non-parametric quantile estimator, QR denotes the semiparametric quantile regression estimator, R-PQR denotes the restricted panel quantile regression estimator, and PQR denotes the unrestricted panel quantile regression estimator. Bootstrap standard errors are shown in parentheses based on 1000 replications. IPW denotes inverse-propensity weighting.*

(3) and (4) of Table 5.1. Although Table 5.1 shows some differences between NP and the QR estimates in column (4), a closer examination of the estimated effects across the 0.05 quantile through the 0.95 quantile reveals that the QTE estimates obtained using QR are similar to the NP estimates.[18] Therefore, the NP results appear to be robust to the use of weights and an alternative parametric specification for estimating the QTE. It is

[18] In estimates not shown here, the NP and QR are statistically significantly different (at the 10-percent level) at 6 quantiles in the interval $\{0.05, 0.06, \ldots, 0.95\}$, which includes the 0.75 quantile shown in Table 5.1 but not the 0.90 quantile. For weighted estimates, the NP and QR are statistically significantly different at 7 quantiles.

important to emphasize that this additional empirical evidence continues to indicate that there is substantial heterogeneity predicted by labor supply theory and low-income women can increase income by reducing hours and claiming welfare, which is consistent so far with the behavioral-induced participation hypothesis.

5.2. **Panel Data Results.** In the last columns of Table 5.1, we present panel quantile regression estimates for both the restricted and unrestricted cases. For the restricted estimator, we let $\lambda_1 = 0.01$ for Jobs First participants and $\lambda_0 = 1$ for AFDC.[19] The weighted R-PQR estimates are shown in column (5) and unweighted estimates in column (6). Despite controlling for women's heterogeneity by treatment status, the R-PQR estimator delivers results that are similar to those of NP and QR at the center of the distribution. In contrast, we observe large differences at the upper quantiles of the conditional distribution of earnings. For the unrestricted case, weighted and unweighted PQR estimates are shown in columns (7) and (8), respectively. In these cases, $\hat{\lambda}$ is estimated to be approximately 0.718 for earnings and 0.673 for total income.[20] We note that the R-PQR and PQR estimates are qualitatively similar. Restricting the degree of shrinkage for individual effects by treatment status imposes no difference at the median, though the unrestricted estimates are somewhat smaller at upper quantiles. Unrestricted penalized estimates, therefore, offer a more conservative contrast to pooled estimates. However, large differences between pooled and panel results in the upper earnings distribution are robust to modeling assumptions on differential shrinkage for individual effects by treatment status. Thus, given the more conservative, yet similar results under weaker assumptions, the unrestricted PQR estimates are preferred as the main results for comparison to the pooled results.

Figure 5.1 compares weighted estimates by NP and PQR for quantiles in the interval from 0.05 to 0.95. The series of estimates appear to be equivalent through the 0.65 quantile after which they tend to diverge at upper quantiles. The reduction of earnings in the upper tail of the distribution was predicted as a natural consequence of behavioral-induced participation attributed to a reduction of exits from welfare. However, when we control for latent individual heterogeneity, the negative treatment effect disappears. Looking at the 0.90 quantile of earnings, for example, there is a weighted NP estimate of -200 dollars compared to a PQR estimate of 300 dollars. As opposed to seeing a negative effect in the upper tail of

[19]The values of λ here are chosen to illustrate the economic model. See Section B.1 in Appendix B for details on tuning parameter selection.

[20]See Appendix B for tuning parameter estimation details, and Section 6.2 for robustness evidence for the tuning parameter selection.

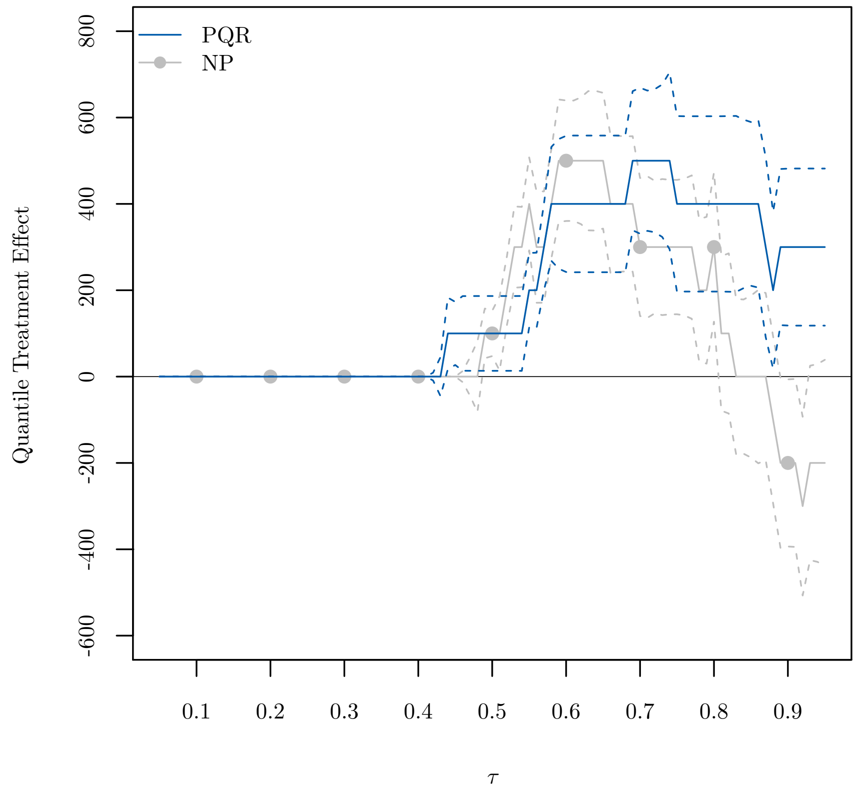

FIGURE 5.1. *Quantile Treatment Effects on the Distribution of Earnings, Quarters 1-7. PQR denotes panel quantile regression estimates and NP denotes non-parametric quantile estimates. The dashed lines represent 90-percent confidence intervals obtained by 1000 bootstrap replications.*

the earnings distribution, the estimated treatment effect continues to be positive and statistically significantly different from zero. Although it is naturally challenging to explain the mechanism behind these differences from the reduced form coefficients, the evidence is consistent with the framework developed in the previous sections, which points to the fact that welfare participation costs can have a differential effect at the upper tail of the conditional distribution of earnings.

5.3. **Welfare Participation.** The individual effect at the 0.90 quantile, $\alpha_i(0.90)$, is intended to capture individual-specific sources of variability, or unobserved heterogeneity that was not adequately controlled for by other covariates. If these latent factors do not affect earnings or are independent of the treatment variable D_{i0}, the proposed panel approach is expected to produce similar findings to other methods. This is not what we observe in Figure 5.1. We interpret the differences between nonparametric estimates and semiparametric panel estimates as suggesting that participation costs of welfare affects Jobs First and AFDC participants differentially at the upper tail. This is consistent with the economic implications discussed in Section 2 since it is expected that high earners who were assigned to AFDC do not participate on welfare and high-earners who were selected to Jobs First do participate. Moreover, it is natural to assume that these women make choices regarding work and welfare based on their opportunity cost of time that depends on family structure, preferences, and program features given treatment status. If there is a nonlinear cost of participation across the distribution of earnings, then controlling for program participation in the pooled model may offer a simple check for the interpretation of labor supply differences that are explained by latent characteristics related to program features.

The experimental data for Connecticut's Jobs First waiver program allow us to run a simple, yet important, robustness check. We have information on whether the individual was receiving cash welfare or Food Stamps at each quarter.[21] Then, if there are latent costs in terms of participation, one can introduce an indicator variable for welfare participation in order to capture the potential omitted variable in the cross-sectional quantile model. We expect, however, small differences in the panel results (PQR) since the method is designed to account for these sources of variability and participation is roughly constant over time.

Let $R_{it} = 1$ if individual i receives either cash welfare or Food Stamps in quarter t, and $R_{it} = 0$ otherwise. In results shown in Figure 5.2, we estimate the QTE by quantile regression by solving

$$\min_{(\beta',\Delta)\in\mathbf{\Theta}} \sum_{i=1}^{N} \sum_{t=1}^{T} \hat{w}_i\left(\mathbf{x}_i\right) \cdot \rho_\tau\left(Y_{it} - \beta_0 - \beta_1 R_{it} - \Delta D_{i0}\right),$$

where $\beta = (\beta_0, \beta_1)'$ and by panel quantile regression by solving

$$\min_{(\beta',\Delta,\alpha')\in\mathbf{\Gamma}} \sum_{i=1}^{N} \sum_{t=1}^{T} \hat{w}_i(\mathbf{x}_i) \cdot \rho_\tau\left(Y_{it} - \beta_0 - \beta_1 R_{it} - \Delta D_{i0} - \alpha_i\right) + \hat{\lambda} \sum_{i=1}^{N} |\alpha_i|,$$

where $w_i(\mathbf{x}_i)$ and λ are estimated as before. Recall that the cross-sectional methods QR and NP offer, as expected, similar point estimates (Table 5.1). Moreover, we found that while the NP (and thus QR) point estimate is equal to -200 dollars at the 0.90 quantile of earnings, PQR suggests a positive treatment effect of 300 dollars (Figure 5.1). It is very interesting to see now that the cross-sectional estimates and panel estimates are roughly equivalent when controlling for participation as in Figure 5.2, suggesting that QR, NP and

[21]For Jobs First, 67.6 percent of women's participation status do not change over quarters 1-7, and for AFDC women, 63.1 percent do not change. The standard deviation for an indicator of participation over this time period is 0.176 in Jobs First and 0.193 in AFDC.

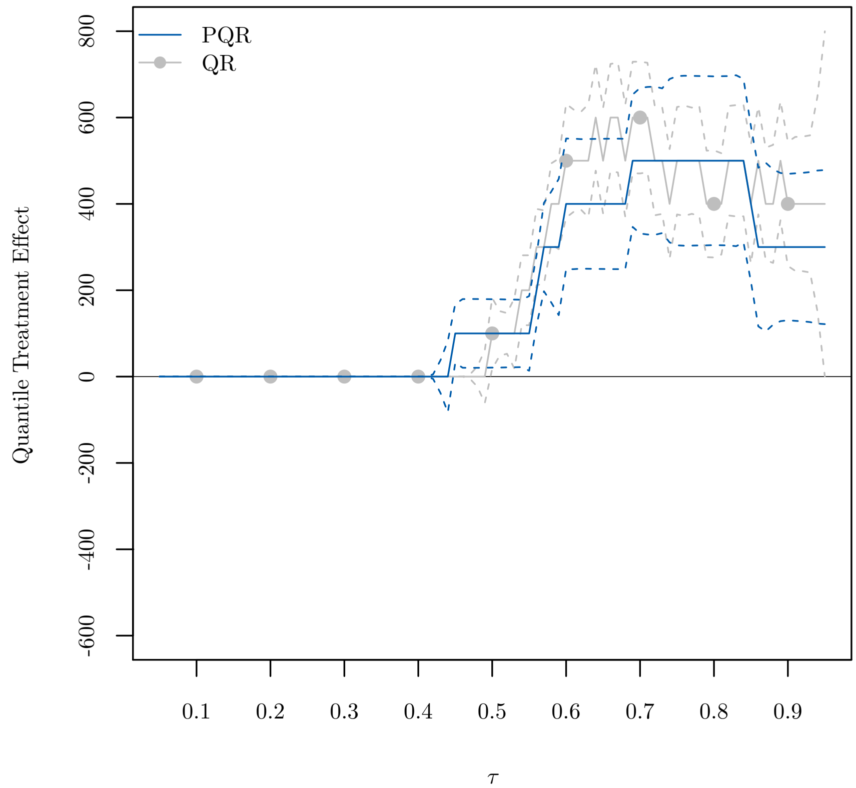

FIGURE 5.2. *Quantile Treatment Effects on the Distribution of Earnings Conditional on Welfare Participation, Quarters 1-7. PQR denotes panel quantile regression estimates and QR denotes quantile regression. The dashed lines represent 90-percent confidence intervals obtained by 1000 bootstrap replications.*

PQR do not offer significantly different results.[22] Also, as expected, the PQR results are robust to the inclusion of a woman's welfare participation status.

5.4. **Quantile Treatment Effects by Participant Type.** The evidence so far suggests that individuals may have differential participation costs by treatment status and that these unobserved costs are increasing in work hours. However, approximately three-fifths of the experimental sample are continuing welfare recipients, whereas the remaining two-fifths of the sample are new applicants. While both participant types would face welfare participation costs such as paperwork and standing in lines, longer-term costs may differ between applicants and recipients.[23] For instance, long-term recipients may have higher informational

[22]The findings of this robustness check are not sensitive to the definition of R_{it}. Results are qualitatively similar for participation defined by cash transfers only, or by Food Stamps only.

[23]Blank and Ruggles (1996) differentiate between participation decisions for women who are persistently eligible versus those who may just qualify for eligibility for a short time.

costs of managing work, child care, and welfare participation because of limited labor market experience. Also, persistent stigma related to long-term welfare participation may affect individuals' beliefs about market productivity.

As noted above, descriptive statistics shown in Table A.1 demonstrate significant differences between samples by participant type. Ongoing recipients are characterized by longer and more frequent welfare spells, as well as higher dependence on public housing and less experience in the labor market. Recipients have less labor force attachment and thus may be affected differentially by informational costs or stigma. Newer applicants, however, are more likely to have higher education and be working more hours at higher wages. The costs of participation for new applicants may be more transitory in nature given that individuals with temporary shocks and better earnings potential may select into welfare under Jobs First based on the generous disregards near the federal poverty line. If there is evidence supporting the behavioral-induced participation hypothesis, it should be related to behavioral responses among applicants as opposed to recipients.

Figure 5.3 shows the QTE by participant type. In panel (a), there is still weak evidence of a negative treatment effect in the upper quantiles for pooled estimates, but the panel estimates are statistically no different from zero throughout nearly the entire distribution of earnings (except at the median). Just as before, there is no difference based on individual effects through the middle of the distribution, and now there is only a small and statistically insignificant difference at the 0.90 quantile. The implication is that participation costs might not play as important of a role for the applicant group as they do for the full sample. Panel (b), however, exhibits large differences in treatment effects at the upper quantiles of the distribution of earnings for recipients, though there is no longer a negative treatment effect for the pooled estimates at the 0.90 quantile. On the other hand, PQR estimates at the upper tail continue to be positive and significant. This evidence suggests that controlling for individual costs of participation matters less for applicants with treatment affects attenuated toward zero and matters more for recipients where the treatment effect is increasing and positive at the upper conditional quantiles.

5.5. **Characterizing Participation Costs using Experimental Data.** The experimental research design allows us to test whether relative participation costs are heterogeneous across the conditional distribution of earnings. A test on the difference between NP and QR is interpreted as a test on the adequacy of the linear parametrization of the model. More importantly, a test on the difference between NP and PQR can be interpreted within the framework discussed in Sections 2 and 3. If participation costs are zero or do not depend on

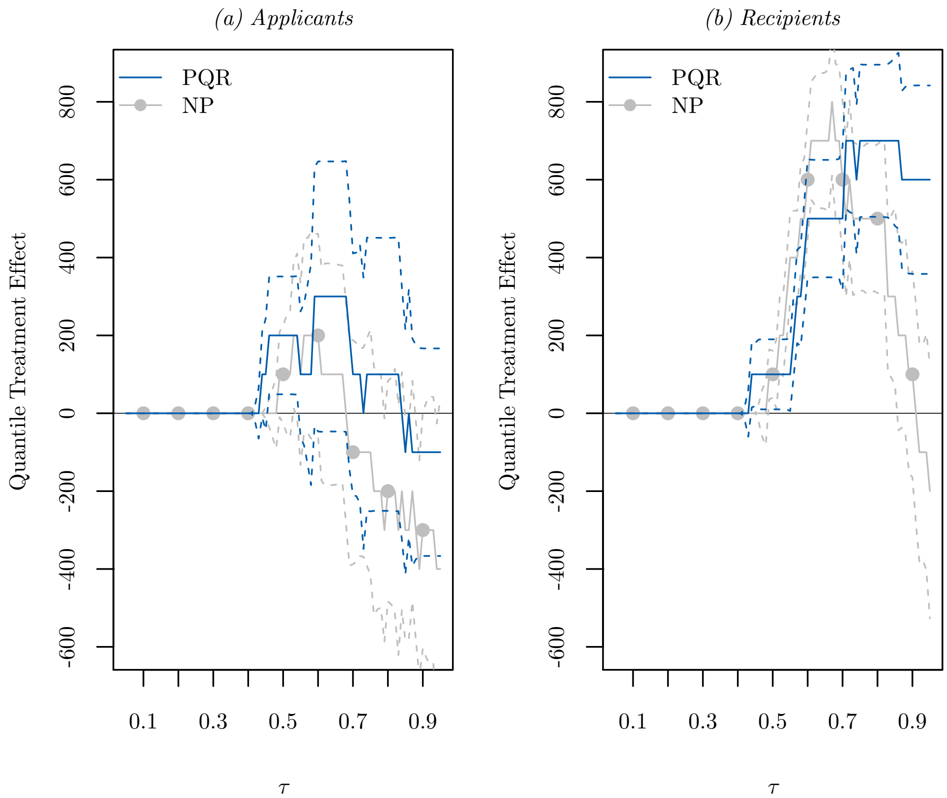

FIGURE 5.3. *Quantile Treatment Effects on the Distribution of Earnings by Participant Type, Quarters 1-7. PQR denotes panel quantile regression estimates and NP denotes non-parametric quantile estimates. The dashed lines represent 90-percent confidence intervals obtained by 1000 bootstrap replications.*

treatment status, then the parameter $C(\tau)$ in equation (3.3) is zero. The framework suggests that $C(0.5) \approx 0$, while $C(0.9) > 0$.

Table 5.2 shows estimator differences at the 0.50 and 0.90 quantiles, with the associated p-values of Hausman-type test statistics as described in Section B.3 (Appendix B). We also show the percentage difference in terms of quarterly earnings and an estimated cost in terms of number of hours per week. As expected, there are no significant differences between the weighted estimates for NP and QR, as shown in column (1). When we turn our attention to weighted estimates by NP and PQR, shown in column (2), we find that there is no statistically significant difference at the 0.50 quantile, and a significant absolute difference of 500 dollars of quarterly earnings at the 0.90 quantile. Columns (3) and (4) show absolute differences between weighted NP and PQR estimates by participant type for applicants and recipients, respectively. For applicants, we fail to reject the null of equality of quantile treatment effects at standard levels, yet for recipients we reject it at the 0.90 quantile at

Quantile τ	Statistic	Full Sample QR−NP (1)	Full Sample PQR−NP (2)	Applicants PQR−NP (3)	Recipients PQR−NP (4)
0.50	Difference	0.00	0.00	100.00	0.00
	p-value	[1.00]	[1.00]	[0.15]	[1.00]
	Percentage Difference	0%	0%	33%	0%
	Hours per week ($6 wage)	0.00	0.00	1.28	0.00
0.90	Difference	0.00	500.00	200.00	500.00
	p-value	[1.00]	[0.00]	[0.34]	[0.01]
	Percentage Difference	0%	10%	4%	11%
	Hours per week ($10 wage)	0.00	3.83	1.53	3.83

Table 5.2. *Estimator Differences for Quantile Treatment Effects on the Distribution of Earnings, Quarters 1-7. NP denotes the non-parametric quantile estimator, QR denotes quantile regression, and PQR denotes panel quantile regression. The p-values shown in brackets are based on 1000 bootstrap replications.*

the 5-percent level with differences similar in magnitude to the full sample estimates. Our findings suggest that Jobs First imposes an estimated cost of 10% of quarterly earnings, and it is larger for recipients than for applicants at high conditional quantiles.[24] In terms of labor supply, the cost of participating in welfare under Jobs First is equivalent to 3.8 less hours per week at the 0.90 quantile of earnings, which is slightly smaller than the estimated AFDC participation cost of about 4 hours per week found by Moffitt (1983, p. 331).

6. Discussion

Although QTE estimates for Jobs First are robust to selection on observables, the panel estimates indicate that the QTE is influenced by unobserved characteristics differentially at the upper conditional quantiles of the earnings distribution. As a possible explanation, we have suggested that women experience nonlinear costs of welfare participation that are increasing with hours worked, which may only be relevant to Jobs First participants who are still eligible at higher earnings. Referring back to the stylized budget constraints in Figure 2.1, a woman's unobserved characteristics might be interpreted as capturing different costs and benefits for welfare participation when labor supply is high. *Ex ante*, labor supply predictions for high earners may be ambiguous conditional on costs of participation given that welfare participants are increasing hours and earnings over time (as documented in

[24]Although the percentage difference for applicants at median earnings is 33%, the difference is statistically insignificant and the magnitude of the percentage is driven by a small denominator of median quarterly earnings.

Figure 4.1) while also being exposed to treatment over time. Based on the QTE estimate differences shown in the previous section, individual heterogeneity plays a prominent role in behavioral responses to Jobs First.

In what follows, this section explores the plausibility of the panel interpretation along with alternative explanations for finding different results only in the upper tail of the earnings distribution.

6.1. **The Pre-Treatment Period and Randomization.** An alternative explanation for finding differences when controlling for individual effects is that randomization may have successfully balanced experimental samples based on observed characteristics though not for unobserved characteristics. To investigate this possibility, it may be helpful to consider earnings trajectories by treatment status before and after randomization, as shown in Figure 6.1.[25] Pre-treatment earnings are mostly censored at the median where as many as 70 percent of the full sample had no earnings 2 years before random assignment. For applicants shown in panel (a), the 75th and 90th percentiles of pre-treatment earnings exhibit a pronounced dip, which is perhaps not surprising in light of the work of Ashenfelter (1978, 1983). However, for recipients shown in panel (b), pre-treatment earnings are relatively flat with no evidence that might suggest induced participation.

If there are important pre-treatment differences in experimental groups, then estimates of the effect of Jobs First on earnings before random assignment may be revealing. Naturally, one would expect zero treatment effect before randomization since the significant earnings disregard had not been implemented yet. Figure 6.2 shows the estimates for the parameter of interest obtained using NP and PQR for pre-treatment earnings, quarters -8 to -1.[26] Through the middle of the pre-treatment earnings distribution, the estimated effects are zero due to the large number of censored observations for earnings before random assignment, thus no information is conveyed about randomization except at the upper quantiles. At the 0.75 quantile of pre-treatment earnings, estimates by the NP estimator and PQR estimator differ by 300 [0.02] while the estimates at the 0.90 quantile only differ by 100 [0.71], with p-values shown in brackets. These results suggest that NP and PQR estimates are similar in the

[25] In the full sample, there are statistically significant differences in earnings by treatment group one year before random assignment, which suggests that differences in pre-treatment shocks may be related to subsequent earnings processes. Meghir and Pistaferri (2004) discuss the importance of unobserved heterogeneity for the variance of earnings related to transitory and permanent income shocks.

[26] The pre-treatment PQR results are based on an estimate of λ equal to 0.716 and pre-treatment earnings are not weighted since predetermined earnings and transfer data are not available.

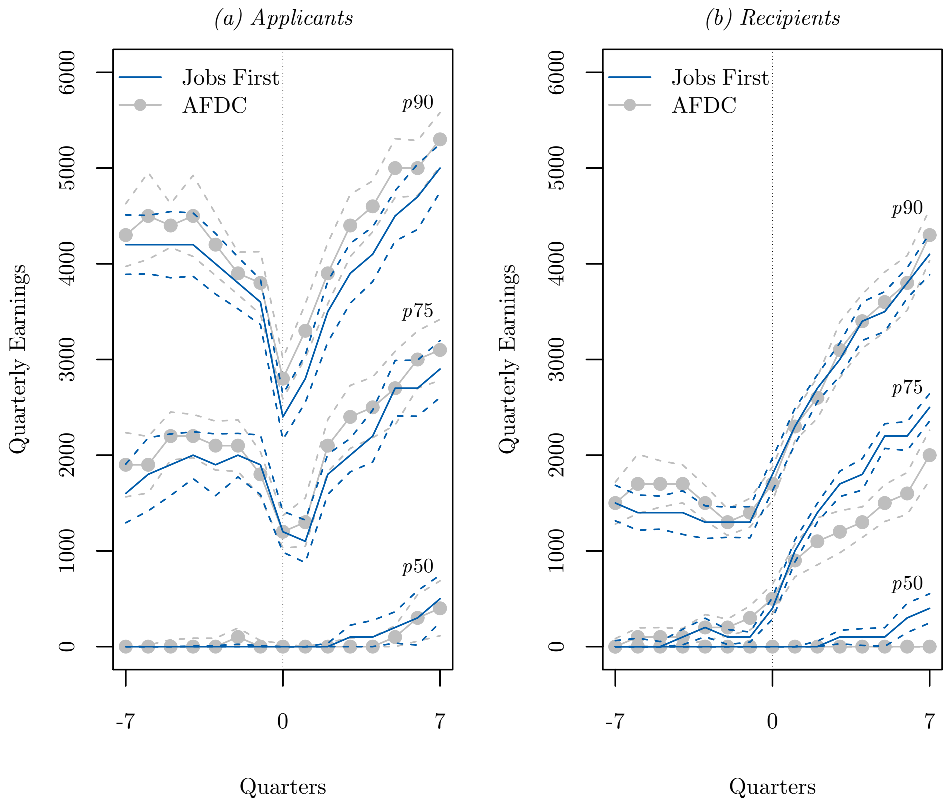

FIGURE 6.1. *Earnings Trajectories by Participant Type and Treatment Group, Quarters -7 to 7. The 50th, 75th, and 95th percentiles of earnings are shown for Jobs First and AFDC before and after random assignment, which is indicated by the dotted line at quarter 0. The dashed lines represent 90-percent confidence intervals obtained by 1000 bootstrap replications.*

pre-treatment period and the direction of effects do not diverge as shown before in Figure 5.1.

6.2. **On the Plausibility of the Sparse Model.** If participation costs, or any other source of latent heterogeneity, are negligible at the upper tail of the earnings distribution, the nonparametric estimator and semiparametric estimator are not expected to produce different results. On the other hand, if costs associated with welfare participation are important in the economic model, one would expect different results because the identifying assumption on observables does not hold against the data, or alternatively, there are non-zero latent costs associated with participation for women who did not exit welfare in the period after the reform. The proposed methodology offers the possibility of investigating these conditions by changing the parameter λ. As the regularization parameter increases, the estimated individual effects $\hat{\alpha}_i(\tau, \lambda)$'s tend to zero, and thus, the PQR estimator converges to QR and

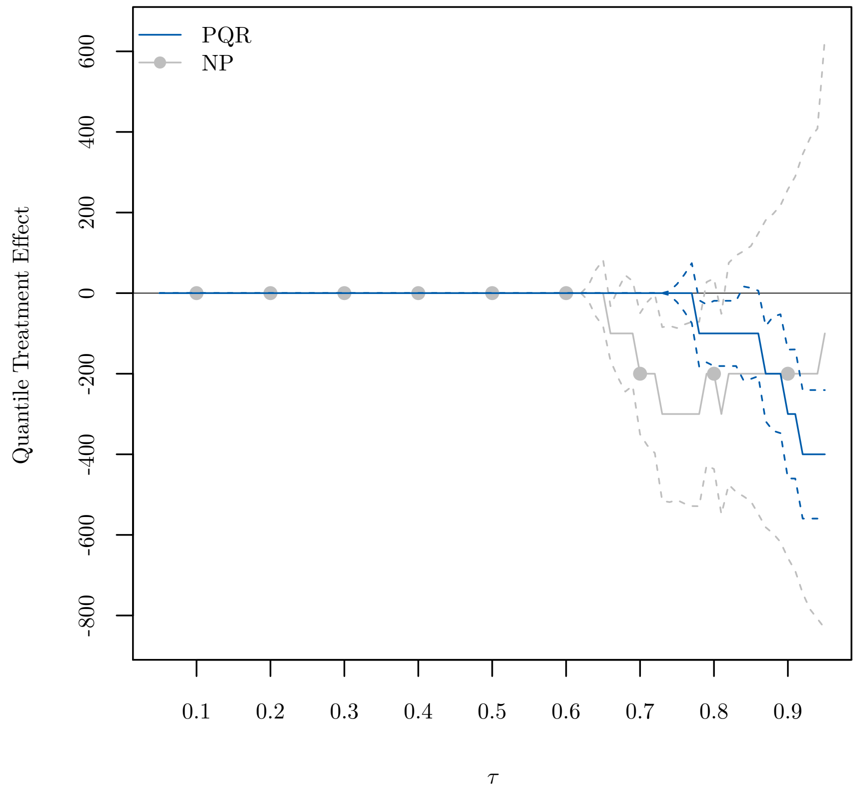

Figure 6.2. *Quantile Treatment Effects on the Distribution of Earnings, Quarters -8 to -1 (Pre-Reform). PQR denotes panel quantile regression estimates and NP denotes non-parametric quantile estimates. The dashed lines represent 90-percent confidence intervals obtained by 1000 bootstrap replications.*

NP. At the same time, unobserved heterogeneity should not affect the QTE estimates for small values of λ by the experimental research design of the program. Recall that costs of participation are assumed to be nonlinear and we expect them to be more pronounced at the upper tail of the earnings distribution. Then, we expect (i) no differences between the PQR and NP estimates at the 0.50 quantile for all values of λ, and (ii) significant differences between PQR and NP at the upper tail for those values of λ that are not sufficiently large enough to shrink the individual effect for all i to zero.

Figure 6.3 shows QTE results for the 0.50 quantile and 0.90 quantile as a function of λ over the interval $(0, 13]$. It also shows the estimated value of λ which is equal to 0.718 for earnings in quarters 1-7. While panel (a) shows results for the median, panel (b) shows results for the 0.90 quantile. As expected, panel (a) demonstrates the case where PQR and NP give similar results. There are no differences between nonparametric estimates and semiparametric panel estimates for all λ, even for $\lambda \to 0$ consistent with a "fixed" effects

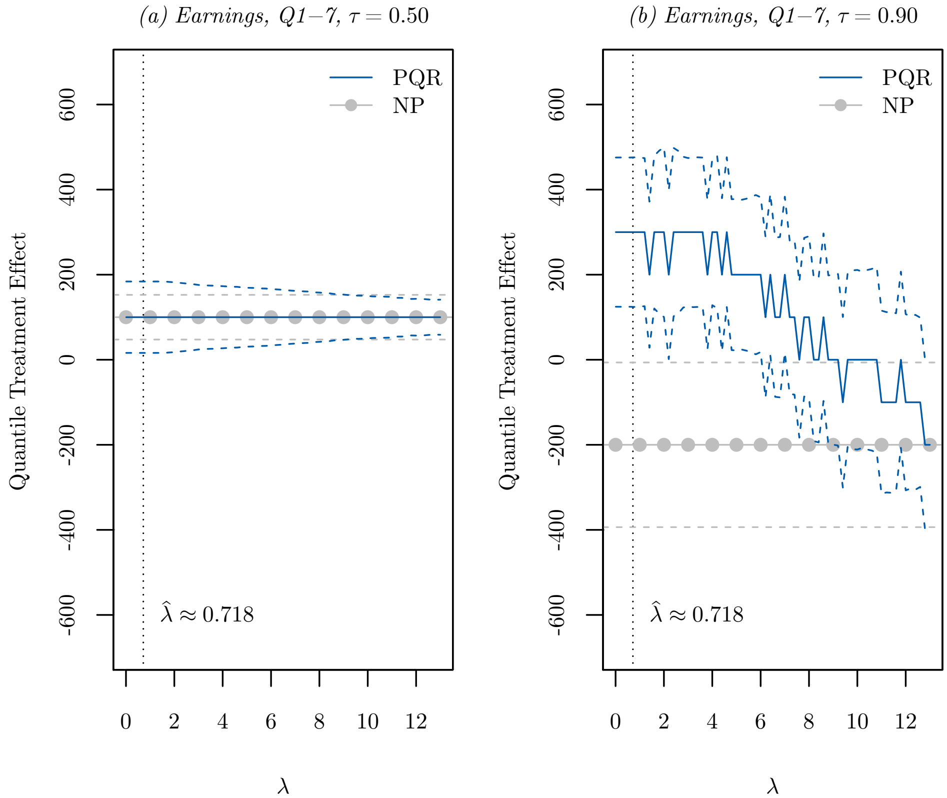

Figure 6.3. *Quantile Treatment Effects on the Distribution of Earnings at the 0.50 and 0.90 Quantiles, Quarters 1-7. PQR denotes panel quantile regression estimates and NP denotes non-parametric quantile estimates. The dashed lines represent 90-percent confidence intervals obtained by 1000 bootstrap replications.*

version of the estimator. On the other hand, panel (b) shows that PQR estimates at the 0.90 quantile can differ substantially from NP estimates. We see that a value of λ less than 4 is consistent with the previous findings since PQR is roughly constant around 300 dollars.

6.3. **Censored Earnings.** When estimating labor supply responses to welfare reform, the considerable amount of censoring at zero earnings may be concerning. To characterize the extent of censoring in the Jobs First data, 70 percent of women had zero earnings two years before random assignment and the average number of censored observations is just above 50 percent for quarters 1-7. Censoring therefore might be expected to affect the shape of the QTE estimates over quantiles τ. We briefly investigated the robustness of the findings to addressing censored observations for earnings and found that censoring is not likely to be driving the empirical results. Censored quantile regression following the methods proposed by Powell (1986) and Fitzenberger (1997), as well as Chernozhukov and Hong (2002), produce similar estimates as QR at the 0.50 quantile and exactly the same estimates at the 0.75 and 0.90 quantiles of the conditional distribution of earnings in quarters 1-7. We argue that the similarity of these results is partially explained by the random assignment of the treatment variable leading to a similar proportion of censored observations by treatment status (49 percent among Jobs First-assigned women and 55 percent among AFDC-assigned women).

6.4. **Policy Relevance of Participation Costs.** Despite the importance of potential work disincentives of welfare generosity, there has been little empirical evidence illustrating behavioral-induced participation where individuals reduce labor supply to gain welfare eligibility. Jobs First has become a primary case study based on the prominent work of Bitler et al. (2006) and Kline and Tartari (2016). The panel estimates of quantile treatment effects shown here should be seen as complementary, though with the important exception of highlighting the role of participation costs. Bitler et al. demonstrated the significance of quantile treatment effects for policy impact analysis, particularly for welfare reform. Our analysis extends their work directly by allowing the costs and benefits of a given program to vary by individual throughout the earnings distribution. Further, we demonstrate that different participant types may incur different costs relative to program exposure as in the case of ongoing recipients compared to new applicants.

Using a structural bounds approach, Kline and Tartari (2016) estimate that the intensive margin effect is bounded by $\{0.28, 1.00\}$ with a 95-percent confidence interval of $[0.20, 1.00]$, which implies that at least 20 percent of women who would have earned above the poverty line are behaviorally induced to reduce hours in order to opt into Jobs First. While Kline and Tartari's model accounts for population heterogeneity by introducing primitives drawn independently from a parametric distribution, our design-based approach can be seen as allowing for earnings to be unconditionally serially dependent. Although our findings are similar in spirit, the evidence presented in this paper suggests that behavioral-induced participation does not generalize to all low-income mothers on welfare near the eligibility threshold, especially regarding continuing recipients, and that a possible explanation of the difference among long- and short-term recipients is participation costs.

7. Conclusion

Behavioral responses to welfare policy, particularly concerning induced participation, are still relevant to public debates for potential reforms regarding TANF and other means-tested transfer programs. It is typically expected however that randomization provides the basis for anticipating that observables and unobservables are equally balanced by treatment status which applies to a range of policy interventions. Motivated by the work of Moffitt (1983) and Blank, Card and Robins (2000), this paper points out the importance of addressing unobserved heterogeneity in the estimation of QTEs using experimental data. We proposed a semi-parametric panel quantile estimator for a model that allows women to vary arbitrarily in preferences and costs of participating in welfare programs. Using data from a welfare reform experiment, we find no evidence of reduced earnings from behavioral-induced participation once we control for unobservables possibly capturing participation costs. The evidence suggests that welfare programs impose different participation costs by treatment status, and these costs can be heterogeneous throughout the conditional earnings distribution.

The literature using observational data had recognized and addressed some of these issues. Moffitt (1983) established the importance of stigma effects, or costs of welfare participation, on labor supply using data from the Panel Study of Income Dynamics. He finds that AFDC participation implies a fixed cost of about four hours of reduced labor supply per week. While Moffitt finds no significant variable stigma, the results are mean estimates under AFDC program rules where eligibility phases out at lower earnings than in the Jobs First context. Regarding other studies related to behavioral-induced participation, the literature has shown little evidence that women reduce hours to opt into welfare given constraints on labor supply adjustments and the lack of bunching near budget constraint notches where welfare phases out (see, e.g., Saez 2010). The literature on distributional effects of welfare reform has mainly abstracted away from the panel nature of policy reform in terms of short-run and long-run effects, which suggests a need to model how low-income single mothers' preferences, incentives and participation costs respond to program features and changes in labor supply over time.

References

Ashenfelter, O. (1978): "Estimating the effect of training programs on earnings," *Review of Economics and Statistics*, 60(1), 47–57.

——— (1983): "Determining participation in income-tested social programs," *Journal of the American Statistical Association*, 78(383), 517–525.

Banerjee, A., R. Hanna, G. Kreindler, and B. A. Olken (2016): "Debunking the stereotype of the lazy welfare recipient: Evidence from cash transfer programs," MIT working paper.

Bargain, O., and K. Orsini (2006): "In-work policies in Europe: Killing two birds with one stone?," *Labour Economics*, 13(6), 667–697.

Becker, G. S. (1965): "A theory of the allocation of time," *Economic Journal*, 75(299), 493–517.

Beffy, M., R. Blundell, A. Bozio, G. Laroque, and M. To (2016): "Labour supply and taxation with restricted choices," Working paper w15/02, Institute for Fiscal Studies.

Belloni, A., and V. Chernozhukov (2011): "ℓ_1-penalized quantile regression in high-dimensional sparse models," *The Annals of Statistics*, 39(1), 82–130.

Belloni, A., V. Chernozhukov, I. Fernández-Val, and C. Hansen (2017): "Program evaluation and causal inference with high-dimensional data," *Econometrica*, 85(1), 233–298.

Bergolo, M., and G. Cruces (2016): "The anatomy of behavioral responses to social assistance when informal employment is high," Dp no. 10197, Institute for the Study of Labor (IZA).

Bitler, M. P., J. B. Gelbach, and H. W. Hoynes (2006): "What mean impacts miss: Distributional effects of welfare reform experiments," *American Economic Review*, 96(4), 988–1012.

Blank, R. M., D. E. Card, and P. K. Robins (2000): "Financial incentives for increasing work and income among low-income families," in *Finding Jobs: Work and Welfare Reform*, ed. by R. M. Blank, and D. E. Card, pp. 373–419. Russell Sage Foundation, New York, NY.

Blank, R. M., and P. Ruggles (1996): "When do women use Aid to Families with Dependent Children and Food Stamps? The dynamics of eligibility versus participation," *Journal of Human Resources*, 31(1), 57–89.

Bloom, D., S. Scrivener, C. Michalopoulos, P. Morris, R. Hendra, D. Adams-Ciardullo, and J. Walter (2002): "Jobs First: Final report on Connecticut's welfare reform initiative," MDRC, New York, NY.

Blundell, R., T. MaCurdy, and C. Meghir (2007): "Labor supply models: Unobserved heterogeneity, nonparticipation and dynamics," *Handbook of Econometrics*, 6, 4667–4775.

Bollinger, C., L. Gonzalez, and J. P. Ziliak (2009): "Welfare reform and the level and composition of income," in *Welfare Reform and Its Long-Term Consequences for America's Poor*, ed. by J. P. Ziliak, pp. 59–103. Cambridge University Press, New York, NY.

Brewer, M., A. Duncan, A. Shephard, and M. J. Suárez (2006): "Did working families' tax credit work? The impact of in-work support on labour supply in Great Britain," *Labour Economics*, 13(6), 699–720.

Burtless, G., and J. A. Hausman (1978): "The effect of taxation on labor supply: Evaluating the Gary negative income tax experiment," *Journal of Political Economy*, 86(6), 1103–1130.

Cattaneo, M. D. (2010): "Efficient semiparametric estimation of multi-valued treatment effects under ignorability," *Journal of Econometrics*, 155(2), 138–154.

Chernozhukov, V., and H. Hong (2002): "Three-step censored quantile regression and extramarital affairs," *Journal of the American Statistical Association*, 97(459), 872–882.

Currie, J. (2006): "The take-up of social benefits," in *Poverty, the Distribution of Income, and Public Policy*, ed. by A. Auerbach, D. Card, and J. Quigley, pp. 80–148. Russell Sage Foundation, New York.

Firpo, S. (2007): "Efficient semiparametric estimation of quantile treatment effects," *Econometrica*, 75(1), 259–276.

Fitzenberger, B. (1997): "A guide to censored quantile regressions," *Handbook of Statistics*, 15, 405–437.

González, L. (2008): "Single mothers, welfare, and incentives to work," *Labour*, 22(3), 447–468.

Gottschalk, P. T. (2005): "Can work alter welfare recipients' beliefs?," *Journal of Policy Analysis and Management*, 24(3), 485–498.

Harding, M., and C. Lamarche (2014): "Estimating and testing a quantile regression model with interactive effects," *Journal of Econometrics*, 178, Part 1, 101 – 113.

Harding, M., and C. Lamarche (2017): "Penalized quantile regression with semiparametric correlated effects: An application with heterogeneous preferences," *Journal of Applied Econometrics*, forthcoming.

Heckman, J. J., H. Ichimura, J. Smith, and P. Todd (1998): "Characterizing selection bias using experimental data," *Econometrica*, 66(5), 1017–1098.

Heckman, J. J., J. Smith, and N. Clements (1997): "Making the most out of programme evaluations and social experiments: Accounting for heterogeneity in programme impacts," *Review of Economic Studies*, 64(4), 487–535.

Imai, K., and M. Ratkovic (2013): "Estimating treatment effect heterogeneity in randomized program evaluation," *Annals of Applied Statistics*, 7(1), 443–470.

Immervoll, H., H. J. Kleven, C. T. Kreiner, and E. Saez (2007): "Welfare reform in European countries: A microsimulation analysis," *Economic Journal*, 117(516), 1–44.

Kline, P., and M. Tartari (2016): "Bounding the labor supply responses to a randomized welfare experiment: A revealed preference approach," *American Economic Review*, 106(4), 972–1014.

Koenker, R. (2004): "Quantile regression for longitudinal data," *Journal of Multivariate Analysis*, 91(1), 74–89.

——— (2005): *Quantile Regression*. Cambridge University Press, New York, NY.

Koenker, R., and G. Bassett Jr (1978): "Regression quantiles," *Econometrica*, 46(1), 33–50.

Lamarche, C. (2010): "Robust penalized quantile regression estimation for panel data," *Journal of Econometrics*, 157(2), 396–408.

Meghir, C., and L. Pistaferri (2004): "Income variance dynamics and heterogeneity," *Econometrica*, 72(1), 1–32.

Meyer, B. D., and D. T. Rosenbaum (2001): "Welfare, the Earned Income Tax Credit, and the labor supply of single mothers," *Quarterly Journal of Economics*, 116(3), 1063–1114.

Moffitt, R. A. (1983): "An economic model of welfare stigma," *American Economic Review*, 73(5), 1023–1035.

——— (2002): "Welfare programs and labor supply," *Handbook of Public Economics*, 4, 2393–2430.

——— (2003): "The Temporary Assistance to Needy Families Program," in *Means-Tested Transfer Programs in the United States*, ed. by R. A. Moffitt, pp. 291–364. University of Chicago Press, Chicago.

Mogstad, M., and C. Ponzato (2012): "Are lone mothers responsive to policy changes? Evidence from a workfare reform in a generous welfare state," *Scandinavian Journal of Economics*, 114(4), 1129–1159.

Powell, J. L. (1986): "Censored regression quantiles," *Journal of Econometrics*, 32(1), 143–155.

Rubin, D. B. (1977): "Assignment to treatment group on the basis of a covariate," *Journal of Educational Statistics*, 2(1), 1–26.

Saez, E. (2010): "Do taxpayers bunch at kink points?," *American Economic Journal: Economic Policy*, 2(3), 180–212.

Słoczyński, T., and J. M. Wooldridge (2016): "A general double robustness result for estimating average treatment effects," *Econometric Theory*, forthcoming.

Ziliak, J. P. (2016): "Temporary Assistance for Needy Families," in *Economics of Means-Tested Transfer Programs in the United States, Volume I*, ed. by R. A. Moffitt, pp. 303–393. National Bureau of Economic Research and University of Chicago Press.

Appendix A. Experimental Sample

This section presents additional descriptive statistics of the experimental sample used in our investigation. To complement the evidence presented in Section 4, Table A.1 presents evidence by participant type.

Appendix B. Technical Appendix

B.1. **Tuning Parameter.** In the empirical analysis of Section 5.2, we estimate the restricted case by setting $\lambda_0 = 1$ for the control group and $\lambda_1 = 0.01$ for Jobs First participants. This restriction imposes values of λ that would directly correspond to model assumptions of individual effects relevant only to the treatment group who are welfare-eligible at higher earnings (and presumably higher hours of labor supply). In this case, the choice of $\lambda_1 = 0.01$ is small enough to allow individual effects that can have a "fixed" effects interpretation for the treatment group, and for the control group, $\lambda_0 = 1$ is set to the standard condition that the variance of α_i is equal to the variance of u_{it} (Koenker 2005, p. 281). For the unrestricted case, λ is determined using a data-driven approach to be approximately 0.718 for the earnings dependent variable and 0.673 for total income. Under the assumption that α_i and D_{i0} are independent, which holds here by experimental design, the tuning parameter λ is estimated to minimize the variance of the QTE estimator (Lamarche 2010). Under further assumptions, λ is equal to the ratio σ_u/σ_α which can be easily estimated by random effects or maximum likelihood methods. Apart from choosing the tuning parameter according to standard values in the literature (as in the restricted case) or by optimizing some objective function (as in the minimum variance estimator in the unrestricted case), another selection criterion could be to follow a grid search over plausible values of λ. This is essentially similar to our robustness results shown for the tuning parameter selection as explored graphically in Figure 6.3.

B.2. **Standard Errors.** We propose to use the bootstrap for inference about $\hat{\Delta}(\tau, \lambda)$. In what follows, for notational simplicity, we suppress the dependence of the QTE estimator on τ and λ. Given the different estimators used in this study, the bootstrap appears to have an advantage over the estimation of the covariance matrices of the limiting process $(NT)^{-1/2}(\hat{\Delta} - \Delta)$. In order to directly compare QTEs with previous estimates, we follow the block bootstrap method used in the panel quantile literature as well as in Bitler et al. (2006). We proceed by drawing a sample with replacement of N subjects including their T observations. Using these new pairs $(\boldsymbol{Y}_i^*, \boldsymbol{D}_i^*, \boldsymbol{x}_i^*)$, we recalculate inverse-propensity weights based on $\boldsymbol{x}_i^*$ in each bootstrap sample, and then obtain Δ^* as the argument that minimizes

Variables	Levels: Recipients	Levels: Applicants	Differences: Unadjusted	Differences: Adjusted	N
Newhaven County (urban)	0.794	0.695	0.099*	0.097*	4803
	(0.404)	(0.460)	(0.013)	(0.013)	
Never married	0.655	0.584	0.072*	0.072*	4803
	(0.475)	(0.493)	(0.014)	(0.015)	
HS dropout	0.350	0.278	0.072*	0.072*	4803
	(0.477)	(0.448)	(0.014)	(0.014)	
More than two children	0.266	0.140	0.125*	0.124*	4803
	(0.442)	(0.348)	(0.011)	(0.013)	
Mother younger than 25	0.237	0.380	-0.144*	-0.147*	4803
	(0.425)	(0.486)	(0.014)	(0.013)	
Mother older than 34	0.323	0.247	0.075*	0.075*	4803
	(0.468)	(0.432)	(0.013)	(0.013)	
Currently working $\geq$ 30 hours	0.253	0.351	-0.098*	-0.092*	989
	(0.435)	(0.478)	(0.030)	(0.029)	
Hourly wage	6.392	7.120	-0.728*	-0.700*	973
	(2.224)	(2.628)	(0.160)	(0.165)	
Public or subsidized housing	0.451	0.191	0.260*	0.260*	4520
	(0.498)	(0.393)	(0.013)	(0.015)	
Ever on AFDC as a child	0.264	0.235	0.030*	0.031*	4491
	(0.441)	(0.424)	(0.013)	(0.013)	
Ever received AFDC at prior quarter 7	0.758	0.196	0.562*	0.562*	4803
	(0.428)	(0.397)	(0.012)	(0.015)	
Length in months of 1st AFDC spell	17.705	13.435	4.269*	4.127*	3607
	(10.332)	(9.980)	(0.345)	(0.362)	
Number of AFDC spells	1.243	1.122	0.121*	0.115*	4803
	(0.595)	(0.689)	(0.019)	(0.020)	
Long-term recipient (> 2 years)	0.730	0.302	0.428*	0.429*	4706
	(0.444)	(0.459)	(0.014)	(0.014)	
Pre-Treatment Quarters					
Average quarterly earnings	413.937	1227.859	-813.922*	-813.649*	4803
	(1041.982)	(1771.295)	(45.170)	(43.484)	
Average quarterly cash welfare	1290.907	197.443	1093.464*	1095.108*	4803
	(659.755)	(462.488)	(16.208)	(25.845)	
Fraction of quarters with earnings	0.259	0.456	-0.197*	-0.199*	4803
	(0.323)	(0.399)	(0.011)	(0.011)	
Fraction of quarters with cash welfare	0.827	0.142	0.684*	0.685*	4803
	(0.309)	(0.296)	(0.009)	(0.014)	
Experimental Quarters 1-7					
Average quarterly earnings	1014.784	1375.760	-360.976*	-351.282*	4803
	(1498.970)	(1781.270)	(49.562)	(53.643)	
Average quarterly cash welfare	1130.330	761.419	368.911*	360.858*	4803
	(607.787)	(617.241)	(18.139)	(20.269)	
Fraction of quarters with earnings	0.466	0.505	-0.039*	-0.044*	4803
	(0.398)	(0.396)	(0.012)	(0.011)	
Fraction of quarters with cash welfare	0.788	0.573	0.215*	0.211*	4803
	(0.319)	(0.393)	(0.011)	(0.011)	

TABLE A.1. *Descriptive Statistics by Participant Type: Applicants and Recipients. Standard deviations are shown in parentheses, and * denotes statistically significant differences at the 10-percent level.*

the objective function. We reiterate this procedure B times to obtain a large sample of realizations $\{\Delta_b^*\}_{b=1}^B$. For a given quantile, we can obtain an estimate of the variance of $\hat{\Delta}(\tau)$ as the sample variance of $\{\Delta_b^*\}_{b=1}^B$. Moreover, a $100(1-2q)\%$ confidence interval can be obtained by constructing the qth quantile and $(1-q)$th quantile of $\{\Delta_b^*\}_{b=1}^B$. This pair bootstrap procedure is applied to compute the estimator standard errors based on $B = 1000$ replications.

B.3. **Hypothesis Testing.** To evaluate significance of differences across quantiles, it is possible to employ the Hausman-type statistic proposed in Harding and Lamarche (2014) or a Wald-type statistic (Koenker 2005). We consider testing a basic general linear hypothesis on a vector $\boldsymbol{\xi}$ of the form $H_0 : \boldsymbol{R}\boldsymbol{\xi} = \boldsymbol{r}$, where $\boldsymbol{R}$ is a matrix that depends on the type of restrictions imposed. For instance, one might evaluate the null hypothesis of equality of effects across quantiles considering a vector $\boldsymbol{\xi} = (\Delta(\tau_1), \ldots, \Delta(\tau_J))'$. More importantly, in Section 5.5 we test for an exogeneity condition of the treatment variable and the independent variables using the Hausman-type test for the null hypothesis that $\Delta(\tau_j)$ in equation (3.1) is equal to $\Delta(\tau_j)$ in equation (3.2) for $j = 1, \ldots, J$.