UKCPR
University of Kentucky Center for Poverty Research

Discussion Paper Series
DP 2014-11

ISSN: 1936-9379

Distributional Effects of Welfare Reform Experiments: A Panel Quantile Regression Examiniation*

Carlos Lamarche
Department of Economics
University of Kentucky

Robert Paul Hartley
Center for Poverty Research
Department of Economics
University of Kentucky

Preferred citation

Lamarche, C., & Hartley, R., Distribution Effects of Welfare Reform Experiments: A Panel Quantile Regression Examination. *University of Kentucky Center for Poverty Research Discussion Paper Series, DP2014-11*. Retrieved [Date] from http://www.ukcpr.org/Publications/DP2014-11.pdf.

Author correspondence

Carlos Lamarche, University of Kentucky, Department of Economics, 301A Mathews Building, Lexington, KY 40506; Email: clamarche@uky.edu; Phone: (859) 257-3371

DISTRIBUTIONAL EFFECTS OF WELFARE REFORM EXPERIMENTS: A PANEL QUANTILE REGRESSION EXAMINATION*

CARLOS LAMARCHE and ROBERT PAUL HARTLEY†

Abstract: In an influential article, Bitler, Gelbach and Hoynes (*American Economic Review*, 2006; 96, 988-1012) illustrate the importance of estimating heterogeneous impacts of welfare reform experiments. They find that the mean treatment effect offers an uninformative summary of opposing effects, while the treatment effects are significantly different across quantiles. We replicate their results and evaluate the robustness of their findings to accounting for individual-specific heterogeneity possibly associated with welfare program participation. We find results that are in general similar to Bitler's et al. findings, although the interpretation of labor supply effects in the upper tail is revised. We find no evidence of behavioral induced participation.

Keywords: Welfare Reform; Quantile Regression; Panel data; Program participation.

JEL Codes: J2, I38, H53, C21, C33.

Distributional effects of policies are increasingly the causal effect of interest among social scientists. In a manuscript published in the September 2006 issue of *American Economic Review*, Bitler, Gelbach and Hoynes (2006) illustrate the importance of estimating quantile treatment effects (QTEs) to analyze a welfare reform experiment. Beginning in 1996, Connecticut implemented a welfare waiver program called Jobs First, under which women experienced more generous earnings disregards compared to Aid to Families with Dependent Children (AFDC). Welfare recipients and applicants were randomized into either Jobs First or AFDC. Bitler et al. find heterogeneous responses in earnings and total income that can be explained using conventional labor supply theory. 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.

*This draft: August 25, 2014. We would like to thank Jim Ziliak for comments and insightful conversations. The data used in this paper are derived from data files made available to researchers by MDRC. The authors remain solely responsible for how the data have been used or interpreted.

†Carlos Lamarche: Department of Economics, University of Kentucky, 302B Mathews Building, Lexington, KY 40506, USA. Email: clamarche@uky.edu. Robert Paul Hartley: Department of Economics, University of Kentucky, 301A Mathews Building, Lexington, KY 40506.

The identification of the quantile treatment effect relied on random assignment data from Connecticut's Jobs First waiver program. As Bitler et al. pointed out, there might be potential issues of selection, which are addressed by using propensity score weighting. Although in principle randomization provides the basis for anticipating that unobservables tend to be similarly balanced across treatment and control, low-income women's preferences over work and welfare may change due to features of either Jobs First or AFDC. It might be possible that interactions of program features and individual characteristics shape preferences differentially as well as heterogeneously with respect to the distribution of earnings. For instance, it is known that behavioral labor supply responses depend on individual costs associated with welfare participation (see, e.g., Moffitt, 1983, Blank, Card and Robins, 2000). An individual who is initially ineligible for welfare will reduce hours and consequently earnings if the additional utility from extra leisure is greater than the utility loss from lower earnings *and* the "stigma" that come with participation. Thus, preferences for working among women assigned to Jobs First might be different than even their own preferences had they been assigned to AFDC and were not on welfare years later. Therefore, when practitioners use repeated measurements of outcome variables in welfare reform experiments, it seems appropriate to model preference heterogeneity possibly associated with program features. Rather than this being a new result, we argue that it is another version of Burtless and Hausman's (1978) argument that two observationally equivalent individuals facing identical budget constraints can respond in very different ways to changes in the constraint. See also Moffitt (1990, 2002) for further discussion on the importance of individual parameters in the stochastic specification of econometric models. For theory and application of distributional analysis for welfare treatment effects, see Blundell, MaCurdy and Meghir (2007), Bollinger, Gonzalez and Ziliak (2009), and Heckman, Smith and Clements (1997), among others.

This paper investigates the robustness of the empirical results in Bitler et al. to addressing latent individual heterogeneity. We first replicate their original results, and then we show that these results can be obtained using standard quantile regression techniques. A parametric quantile regression estimator produces roughly the same point estimates as a non-parametric estimator. This result is interpreted as evidence suggesting that the treatment has a linear effect. In addition, we repeat the analysis without using inverse-propensity score weighting and we find that the Bitler et al. conclusions are qualitatively robust to the unweighted estimation. In this paper, however, we find that accounting for unobservable heterogeneity, possibly associated with individual costs of welfare participation, matters for understanding responses of low-income women at the upper tail of the conditional distribution of the response variable. While the results are generally similar to Bitler's et al. results, the interpretation of a low-income woman's labor supply in the upper tail is revised. We find no evidence of reduced earnings from behavioral induced participation.

1. Empirical Approach

Consider 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 status, respectively. We then write $Y_{it} = D_{it}Y_{1,it} + (1 - D_{it})Y_{0,it}$. Lastly, let $Q_Y(\tau)$ denote the τ-th quantile of the distribution of Y.

Bitler et al. motivate their empirical results using predictions from a static labor supply model and find evidence that the QTE, $\Delta(\tau) = Q_{Y_1}(\tau) - Q_{Y_0}(\tau)$, is not constant across quantiles τ. As shown in Section 3.2, their findings can be obtained by estimating the QTE using the following linear equation: $Y_{it} = Y_{0,it} + \Delta D_{it}$. Consider now that the treatment is subject-specific satisfying $\Delta_i + Y_{0,it} = Y_{1,it}$. This can be motivated by changes in preferences given the specific features of the policy reform, fundamentally reflecting that a low-income woman can respond in different ways to welfare participation. Under the simplifying assumption that $\Delta_i = v_i + \Delta$, it follows 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 time invariant, $Y_{it} = Y_{0,it} + \alpha_i + \Delta D_i$ where $\alpha_i = v_i D_i$. Note that this leads to a sparse model where α_i is equal to zero if $D_i = 0$ and $\alpha_i = v_i$ if $D_i = 1$.

At the upper quantiles of earnings, the model has a natural interpretation. Consider an earnings level that exceeds the sum of the poverty line and the maximum benefit (i.e., point H, Figure 1, Bitler et al. 2006). The location assigned to AFDC is above the poverty line, and consequently, there are no costs of welfare participation (e.g., $\alpha_i = 0$). On the other hand, woman i can choose to reduce hours of work without a reduction in total income. The reduction in earnings can be compensated with transfers. In this case, there is also an individual cost of welfare participation, here labeled $\alpha_i = v_i$. Therefore, there is a range of earnings toward the upper quantiles where behavioral induced eligibility effects might not be observed if the program evaluator accounts for welfare participation costs, including stigma.

The QTE, $\Delta(\tau)$, can be estimated by using panel data quantile regression methods (see, e.g., Koenker, 2004; Lamarche, 2010). The estimator $\hat{\Delta}(\tau, \lambda)$ is obtained from a model that controls for latent individual heterogeneity and is a function of a tuning parameter λ. Under independence between v_i and D_i, which holds here by randomization, this estimator produces more precise estimates while controlling for unobserved heterogeneity. It can also ameliorate potential issues associated with incidental parameters in more general models than the one estimated in this paper. In addition, the advantage of the method in this setting is that it can estimate sparse models and control for a low-income woman's latent heterogeneity while identifying the effect of the time-invariant treatment indicator, D_i.

2. Data and Welfare Reform

The data used in this comment were obtained from the AER website and MDRC (formerly Manpower Demonstration Research Corporation). The experimental evaluation of Connecticut's Jobs First waiver by MDRC was used by Bitler et al. (2006). Table A.1 in the Appendix reproduces Table 3 in Bitler et al. (2006) offering sample means for several demographic characteristics, earnings, cash welfare and food stamps among women in the experimental data. Although we use data from a randomized welfare experiment, the means of a few variables are statistically significantly different by treatment status. This leads to Bitler's et al. approach of employing inverse-propensity score weighting to control for selection on observables, though this does not influence results qualitatively.

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 Connecticut Department of Social Services along with MDRC collected data for 7 to 8 months before random assignment and at least 16 months after. The data available represent a panel of earnings, transfers, and total income for 4803 women. Transfers is the sum of cash welfare and food stamps, and total income is the sum of earnings and transfers.

The key feature of this waiver program is a 100-percent earnings disregard (up to the federal poverty line), which leads to an implicit tax of zero percent. In contrast, AFDC disregarded the first \$120 of monthly earnings for the first year in the program and \$90 after the first year. Other key differences are time limits and sanctions. While Jobs First has a strict 21-month time limit, AFDC has no time limits. See Table 1 in Bitler et al. (2006) as well as Bloom et al. (2002) for a detailed description of differences between Jobs First and AFDC programs.

3. Empirical results

3.1. **Replication and weighted quantile treatment effects.** Bitler et al. (2006) estimate the QTE by measuring the difference between the empirical cumulative distributions of the treatment and control groups for each variable of interest: earnings, transfers, and total income. Their QTE is simply the difference between the τ^{th} quantile of the response

Quantile τ	BGH (1)	BGH (2)	QR (3)	QR (4)	PQR (5)	PQR (6)
			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)
0.25	0.00	0.00	0.00	0.00	0.00	0.00
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
0.50	100.00	100.00	100.00	100.00	200.00	100.00
	(30.05)	(28.24)	(35.07)	(23.34)	(49.56)	(48.24)
0.75	300.00	300.00	300.00	100.00	576.23	500.00
	(90.28)	(123.98)	(100.90)	(101.18)	(77.15)	(97.75)
0.90	-200.00	-200.00	-200.00	-300.00	442.02	200.00
	(115.30)	(212.31)	(124.50)	(123.77)	(87.75)	(83.95)
			Total Income, Quarters 8-16			
0.10	0.00	0.00	-0.00	0.00	-286.40	-400.00
	(0.00)	(0.00)	(0.00)	(0.00)	(75.28)	(72.02)
0.25	-150.00	-150.00	-150.00	-150.00	-25.94	-150.00
	(107.20)	(113.74)	(117.05)	(113.08)	(42.24)	(48.36)
0.50	50.00	50.00	50.00	50.00	170.36	50.00
	(66.38)	(71.74)	(68.39)	(68.12)	(55.18)	(55.64)
0.75	300.00	250.00	300.00	200.00	295.28	250.00
	(89.31)	(122.13)	(96.09)	(97.47)	(58.41)	(67.23)
0.90	0.00	0.00	0.00	0.00	449.61	328.00
	(111.27)	(210.86)	(120.58)	(115.05)	(75.89)	(64.34)
IPS weights	Yes	No	Yes	No	Yes	No
Individual effects	No	No	No	No	Yes	Yes

Table 3.1. *Quantile Treatment Effects on the Distribution of Earnings (Quarters 1-7) and Total Income (Quarters 8-16). BGH denotes nonparametric quantile estimates as in Bitler et al. (2006), QR denotes quantile regression and PQR denotes panel quantile regression. Bootstrap standard errors are shown in parentheses based on 1000 replications. IPS denotes inverse-propensity score weighting.*

variable for women in the Jobs First program and the τ^{th} quantile of the response variable for women in the AFDC program. In what follows, we refer to QTE as a general parameter of interest and label their approach BGH. 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 clearly evident before time limits apply to women assigned to Jobs First.

Table 3.1 presents results for the QTE estimated as in Bitler et al. (2006). These results are compared to estimates obtained from quantile regression (QR) and panel quantile regression for a model with individual effects (PQR). It is worth noting that we reproduce all of the BGH estimates exactly with only very slight variations in the confidence intervals for these estimates based on different random samples used for the 1000 bootstrap replications.[1] Following Bitler et al., we use inverse-propensity score weighting to obtain the QTE estimates shown in column (1); estimates obtained without weights are shown in column (2). 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. As the authors mention, and we confirm in our replication, 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 BGH estimates is at the 0.75 quantile for total income in quarters 8-16: 300 in column (1) and 250 in column (2). The data are measured in discrete units with dollars rounded to the nearest hundred, so a quantile estimate difference of 50 may represent a small difference. 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.

3.2. **Quantile regression.** Under the assumption that the treatment effect is linear and treatment status is randomly assigned, the estimator for the QTE used in Bitler et al. and the quantile regression estimator for a model that conditions on the treatment indicator variable are expected to yield similar results (see, e.g., Koenker, 2005). The weighted and unweighted QR results are shown in columns (3) and (4) of Table 3.1. The results for QR are similar to those of BGH.

Table 3.1 shows some differences between BGH and the QR estimates in column (4), although a closer examination of the estimated effects across 91 equally spaced quantiles

[1]Of the 4803 women in the sample, 30 are missing data for quarter 16. Bitler et al. note that estimation for quarters 8-16 uses $4773 \times 9 = 42957$ individual-quarter observations; however, their code obtained from the AER website and figures are consistent with the use of $4773 \times 9 + 30 \times 8 = 43197$ observations. We include all 4803 women and reproduce their estimates for total income in quarters 8-16.

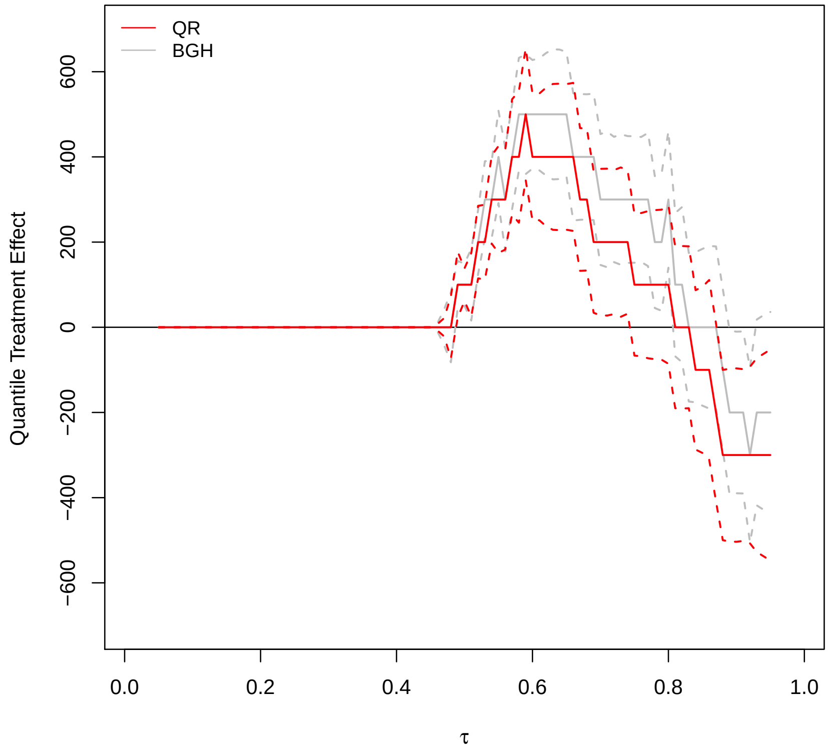

FIGURE 3.1. *Quantile Treatment Effects on the Distribution of Earnings, Quarters 1-7. BGH shows estimates obtained as in Bitler et al. and QR denotes quantile regression estimates obtained without inverse-propensity score weighting. The dashed lines represent 90-percent confidence intervals obtained by 1000 bootstrap replications.*

reveals that the estimated QTE estimates obtained using QR are similar to the BGH estimates. Figure 3.1 provides a graphical comparison of BGH and the unweighted QR where the similarity of estimates may be more apparent. Therefore, the BGH results appear to be robust to the use of weights and an alternative parametric specification for estimating the QTE. It is 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 with the behavioral induced participation hypothesis.

3.3. **Panel quantile regression.** As discussed in the previous sections, it is of fundamental importance to allow for individual parameters in the stochastic specification of QTEs. In the last columns of Table 3.1, we present the weighted (column 5) and unweighted (column

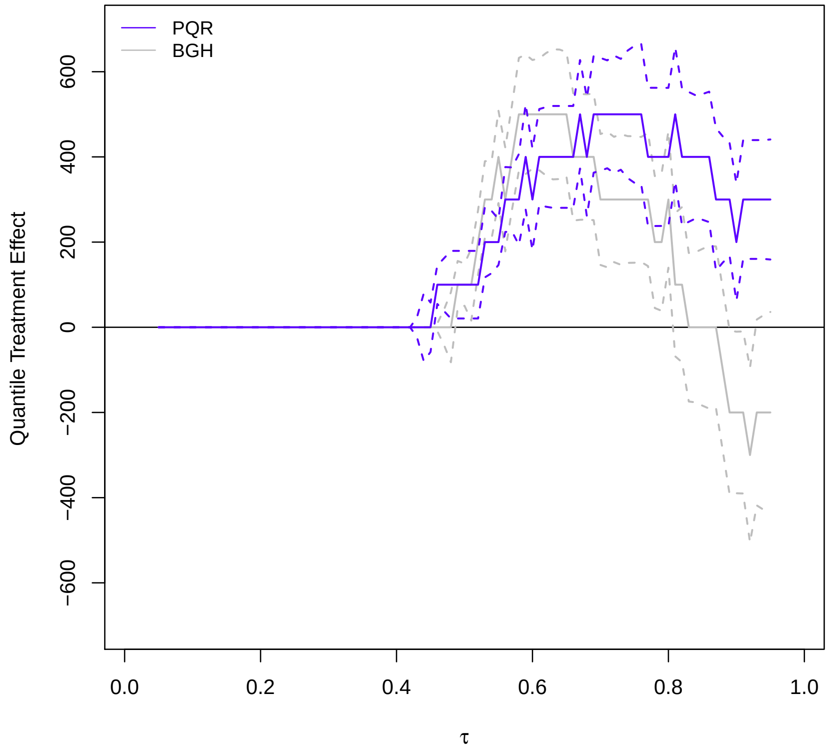

FIGURE 3.2. *Quantile Treatment Effects on the Distribution of Earnings, Quarters 1-7. BGH shows estimates obtained as in Bitler et al. and PQR denotes panel quantile regression estimates. The dashed lines represent 90-percent confidence intervals obtained by 1000 bootstrap replications.*

6) panel quantile regression results. The PQR results were obtained by estimating λ as suggested in the literature (Koenker 2005, Lamarche 2010). We estimate λ to be approximately 0.718 for earnings and 0.673 for total income. In spite of controlling for individual heterogeneity, the PQR estimator delivers results that are similar to those offered by BGH and QR at the center of the distribution. In contrast, we observe significant differences at the upper quantiles of the conditional distribution of earnings.

One of the interesting results in Bitler et al. is the reduction of earnings for women in the upper tail of the distribution. This response was predicted as a natural consequence of a 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 an unweighted BGH estimate of -200 compared with a PQR estimate of 200. As opposed to seeing a negative effect in

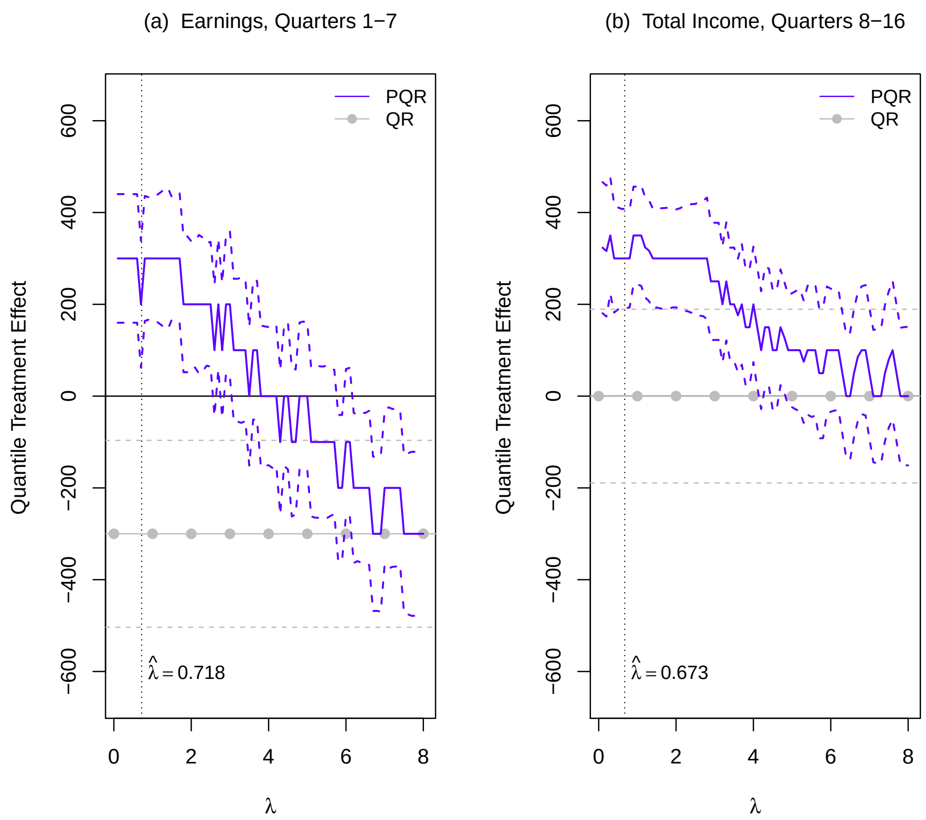

FIGURE 3.3. *Quantile Treatment Effects at the 90th Quantile of the Distribution of (a) Earnings, Quarters 1-7, and (b) Total Income, Quarters 8-16. PQR shows panel quantile regression estimates and QR denotes quantile regression estimates. The dashed lines represent 90-percent confidence intervals obtained by 1000 bootstrap replications.*

the upper tail of the earnings distribution, the estimated treatment effect continues to be positive and statistically significantly different than zero (Figure 3.2).

Figure 3.3 shows QTE results for the 0.90 quantile as a function of λ in order to emphasize the differences when accounting for unobserved heterogeneity in the upper tail of the earnings and total income distributions. These panels demonstrate the importance of the choice of λ to the robustness of PQR results. A value of λ near zero gives results similar to the QTE obtained from a panel quantile model with fixed effects, whereas a parameter value approaching infinity produces QTE results that are equivalent to QR. Each panel shows PQR results at the 0.90 quantile for 80 equally spaced λ's defined over the interval $(0, 8]$. We find that a value of λ less than 2 is consistent with the limiting case of fixed effects quantile regression results since the QTE estimated by PQR is roughly constant around 200 to 300 dollars. Panel (a) also shows that the QTE point estimates obtained by PQR converge to the QR results as λ gets near 8, although the PQR point estimates are more precise. Panel (b) shows that there are significant differences at the upper tail of total income, and the effect is not negligible when we account for latent heterogeneity.

4. Conclusion

It is typically expected that randomization would provide the basis for anticipating that observables and unobservables are equally balanced by treatment status. This naturally applies to the analysis of welfare reform experiments. 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 QTE using experimental data. We revisited one of the most influential empirical studies of distributional effects of welfare reforms, Bitler, Gelbach and Hoynes (2006), and find that their conclusions are qualitatively robust to selection bias and they are generally similar to results obtained by other quantile techniques. In models with individual effects, however, we find no evidence of reduced earnings from behavioral induced participation.

References

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.

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.

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.

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.

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.

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

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

——— (1990): "The econometrics of kinked budget constraints," *Journal of Economic Perspectives*, 4(2), 119–139.

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

Appendix A. Experimental data

	Levels		Differences	
Variables	Jobs First	AFDC	Unadjusted	Adjusted
White	0.362	0.348	0.014	0.001
Black	0.368	0.371	-0.003	-0.000
Hispanic	0.207	0.216	-0.009	-0.001
Never married	0.654	0.661	-0.007	-0.000
Div/wid/sep/living apart	0.332	0.327	0.005	0.000
HS dropout	0.350	0.334	0.017	-0.000
HS diploma/GED	0.583	0.604	-0.021	0.001
More than HS diploma	0.066	0.062	0.004	0.000
More than two children	0.235	0.214	0.021*	-0.000
Mother younger than 25	0.289	0.297	-0.007	-0.000
Mother age 25-34	0.410	0.418	-0.007	0.000
Mother older than 34	0.301	0.286	0.015	0.000
Recipient (stock) sample	0.624	0.593	0.031*	-0.001
Earnings	678.908	785.895	-106.988*	-0.887
Cash welfare	890.818	835.112	55.706*	-0.833
Food stamps	352.117	339.352	12.764	0.316
Any earnings	0.322	0.351	-0.029*	0.000
Any cash welfare	0.573	0.544	0.029*	-0.001
Any food stamps	0.607	0.598	0.009	0.000

Table A.1. *Descriptive Statistics by Treatment Status. The table reports sample averages and* * *denotes statistically significantly different at 10 percent.*