American Economic Review 2016, 106(4): 972–1014
http://dx.doi.org/10.1257/aer.20130824

Bounding the Labor Supply Responses to a Randomized Welfare Experiment: A Revealed Preference Approach†

By Patrick Kline and Melissa Tartari*

We study the short-term impact of Connecticut's Jobs First welfare reform experiment on women's labor supply and welfare participation decisions. A nonparametric optimizing model is shown to restrict the set of counterfactual choices compatible with each woman's actual choice. These revealed preference restrictions yield informative bounds on the frequency of several intensive and extensive margin responses to the experiment. We find that welfare reform induced many women to work but led some others to reduce their earnings in order to receive assistance. The bounds on this latter "opt-in" effect imply that intensive margin labor supply responses are nontrivial. (*JEL* H23, H75, I38, J16, J22)

The United States, like other advanced economies, has an extensive system of transfer programs designed to provide social insurance and improve equity. By affecting work incentives, these programs can induce individuals to enter or exit the labor force (extensive margin responses) or to alter how much they earn conditional on working (intensive margin responses). The relative magnitude of these responses is an important input to the optimal design of tax and transfer schemes (Diamond 1980; Saez 2002; Laroque 2005).

Much of the empirical literature concludes that adjustment to policy reforms occurs primarily along the extensive margin.[1] Two sorts of evidence are often cited in support of this position. First, several studies exploiting policy variation fail to find evidence of mean impacts on hours worked among the employed (Eissa and Liebman 1996; Meyer and Rosenbaum 2001; Meyer 2002). Second, in both survey

* Kline: UC Berkeley, 530 Evans Hall 3880, Berkeley, CA 94720 (e-mail: pkline@econ.berkeley.edu); Tartari: University of Chicago, 5757 South University Avenue, SHFE 012, Chicago, IL 60637 (e-mail: mtartari@uchicago.edu). This paper previously circulated under the title "What Distributional Impacts Mean: Welfare Reform Experiments and Competing Margins of Adjustment." We thank Andres Santos and seminar participants and discussants at the NBER Public Economics and Labor Studies Program Meetings, the Becker-Friedman Institute Interactions Conference, the Institute for Research on Poverty meetings, the Federal Reserve Bank of Chicago, the Harris School, Mannheim University, MIT, UC Berkeley, UCL, UCLA, UCSD, Rice University, the University of Sydney, the University of Technology Sydney, Stanford, and Yale for useful comments. Stuart Craig, Attila Lindner, and Raffaele Saggio provided outstanding research assistance. This research was supported by NSF grant 0962352.

† Go to http://dx.doi.org/10.1257/aer.20130824 to visit the article page for additional materials and author disclosure statement(s).

[1] Heckman (1993, p. 118), for instance, concludes that "elasticities are closer to zero than one for hours-of-work equations (or weeks-of-work equations) *estimated for those who are working*. A major lesson of the past 20 years is that the strongest empirical effects of wages and nonlabor income on labor supply are to be found at the extensive margin" (emphasis in original). Likewise, many modern models of aggregate labor supply are now predicated on the notion that labor supply is "indivisible" (Hansen 1985; Rogerson 1988; Ljungqvist and Sargent 2011). See Chetty, Guren, Manoli, and Weber (2011) for an assessment of how macro estimates of these models compare to estimates from micro data.

and administrative data, earnings tend not to exhibit much bunching at the budget "kinks" induced by tax and transfer policies, suggesting that intensive margin elasticities are small (Heckman 1983; Saez 2010). Both forms of evidence are subject to qualification. In addition to being susceptible to sample selection bias, mean impacts on hours worked ignore the potentially offsetting labor supply effects of program phase-in and phase-out provisions (Bitler, Gelbach, and Hoynes 2006). And although excess mass at kink-points is a nonparametric indicator of intensive margin responsiveness (Saez 2010), demand-side constraints on labor supply may confound the quantitative inferences drawn from bunching approaches (Chetty, Friedman, Olsen, and Pistaferri 2011).

This paper studies the impact of Connecticut's Jobs First (JF) welfare reform experiment on the labor supply and program participation decisions of a sample of welfare applicants and recipients. We develop a nonparametric approach to measuring intensive and extensive margin responses to the JF reform that remains valid in the presence of labor supply constraints, impact heterogeneity, and self-selection. Conceptually, detecting adjustment along a given margin in response to a policy reform requires inferring what choices a decision maker would have made if the reform had not taken place. Because choices are only observed under the policy regime to which the decision maker is exposed, the problem of distinguishing response margins is closely tied to fundamental challenges in causal inference. To address these challenges, we follow Manski (2014) in using revealed preference arguments to restrict the set of counterfactual choices compatible with each decision maker's actual choice. These restrictions are shown to yield informative bounds on the frequency of intensive and extensive margin responses to reform when policy regimes are randomly assigned.

The JF experiment provides an interesting venue for studying labor supply because the reform entailed a mix of positive and negative work incentives characteristic of many transfer programs. First, it strengthened work requirements and increased sanctions for welfare recipients who fail to seek work. Second, it changed the manner in which welfare benefits phase out by disregarding earnings up to an eligibility threshold (or "notch") above which benefits abruptly drop to zero. Bitler, Gelbach, and Hoynes—henceforth, BGH—show that the JF reform induced a nuanced pattern of quantile treatment effects (QTE) on earnings qualitatively consistent with intensive margin responsiveness (BGH 2006). They find that JF boosted the middle quantiles of earnings while lowering the top quantiles, yielding a mean earnings effect near zero. The negative impacts on upper quantiles provide suggestive evidence of an "opt-in" response to welfare (Ashenfelter 1983), whereby working women are induced to lower their earnings in order to qualify for transfers.

Quantifying the frequency of intensive and extensive margin responses to this reform requires additional structure, as the experiment may have shifted women between many points in the earnings distribution, potentially violating the standard "rank preservation" condition needed to infer impact distributions from QTEs.[2]

[2] Here rank preservation means that the JF reform would not alter a woman's rank in the distribution of earnings. With this restriction, QTEs can be used to infer the joint distribution of potential earnings under the two policy regimes (Heckman, Smith, and Clements 1997) and hence to quantify extensive and intensive margin labor supply responses. BGH are skeptical of the rank preservation assumption and, in a related analysis, provide evidence of rank reversals in the Canadian Self-Sufficiency Project experiment (BGH 2005). In the JF experiment, rank

To narrow down the set of possible responses to the experiment, we develop a nonparametric optimizing model of labor supply and welfare participation. In the model, women value consumption and may derive disutility from welfare participation and work. Labor supply decisions are potentially constrained by the set of job offers drawn, and earnings can, at some cost, be underreported to the welfare agency, which explains the empirical finding that some women with earnings above the eligibility notch draw welfare benefits. Despite this generality, the model places sharp restrictions on how women may respond to the JF reform that follow from simple revealed preference arguments. Specifically, if the utility of a woman's choice under existing rules was not lowered by the reform, she will either make the same choice under JF or select an alternative that the reform made more attractive.

In taking the model to the data, we permit women to vary arbitrarily in their preferences and constraints, which may also evolve over time in an unrestricted fashion. This flexibility allows us to rationalize any distribution of earnings and program participation choices found under a given policy regime.[3] Nevertheless, we show that our model places strong testable restrictions on the experimental impacts generated by the JF reform. Specifically, we use the aforementioned revealed preference restrictions to develop analytic bounds on the proportion of women responding along each of nine allowable margins defined by pairings of coarse earnings and program participation categories across policy regimes.

Applying our identification results, we find evidence of substantial intensive and extensive margin responses to reform over the first seven quarters of the experiment. The JF reform incentivized at least 14 percent of the women who would not have worked to do so and more than 32 percent of women who would have worked off welfare at low earnings to take up assistance. Importantly, we find that at least 20 percent of the women who would have worked off welfare at relatively high earnings levels were induced to reduce their earnings and opt in to welfare, demonstrating that the reform in fact led to substantial intensive margin responses. We also find that the JF work requirements induced at least 2 percent of the women who would have not worked while on welfare to work and underreport their earnings in order to maintain eligibility for benefits.

Our results demonstrate that simple revealed preference arguments allow researchers studying policy reforms to derive informative bounds on the size of competing response margins under very weak assumptions. These findings extend results by Heckman, Smith, and Clements (1997) who, in the context of an application to the US Job Training Partnership Act, considered the identifying power of Roy (1951)-type models of optimization for the joint distribution of potential outcomes. Our approach is applicable to more general settings that do not obey strong Roy-style selection on potential outcome differences, and can easily be adapted to other reforms which alter the value of alternatives in known directions.

reversals could occur (for example) because reform induced some skilled women to work and earn above the eligibility notch while leading others to lower their earnings below the notch through an opt-in response.

[3] This is in contrast to traditional parametric models of labor supply (e.g., Burtless and Hausman 1978; Hoynes 1996; Keane and Moffitt 1998) that can be identified without policy variation. See MaCurdy, Green, and Paarsch (1990) for an early critique of parametrically structured econometric models of labor supply with nonlinear budget sets.

We also contribute to a recent literature on partial identification of labor supply models. The bounding approach developed here is closely related to the theoretical analysis of Manski (2014), who considers the use of revealed preference arguments to set-identify tax policy counterfactuals. While Manski conducts computational experiments involving a single tax parameter, we study a reform that changes a bundle of policy features and employ a correspondingly richer model incorporating program participation and reporting decisions. Our additional allowances for labor supply constraints and endogenous wages necessitate the use of policy variation to achieve identification. Blundell, Bozio, and Laroque (2011a,b) also implement a bounds based analysis of labor supply behavior but are concerned with a statistical decomposition of fluctuations in aggregate hours worked rather than formal identification of policy counterfactuals. Their findings, which are compatible with ours, indicate that adjustments along both the intensive and extensive margins are important contributors to fluctuations in aggregate hours worked. Finally, Chetty (2012) considers bounds on labor supply elasticities in a class of semiparametric models with optimization frictions. He too finds evidence of nontrivial intensive margin responsiveness, but relies on strong parametric assumptions.

The remainder of the paper is structured as follows. Section I describes the Jobs First experiment. Section II describes the data from the Jobs First Public Use Files. Section III summarizes the program's experimental impacts on earnings and provides a test for intensive margin responsiveness. Section IV describes our baseline optimizing model. Section V derives the restrictions implied by revealed preference. Section VI considers some extensions of the baseline model. Section VII studies identification of the probabilities of responding to reform along various margins. Section VIII discusses the computation of bounds on response probabilities and inference issues. Section IX provides our main empirical results and Section X discusses their robustness. Section XI concludes. Technical proofs and additional results are provided in an online Appendix.

I. The Jobs First Evaluation

With the passage of the Personal Responsibility and Work Opportunity Reconciliation Act in 1996, all 50 states were required to reform their Aid to Families with Dependent Children (AFDC) welfare programs by introducing lifetime time limits, work requirements, and enhanced financial incentives to work while on assistance. The state of Connecticut responded to these changes by implementing the Jobs First program. To study the effectiveness of the reform, the state contracted with the Manpower Development Research Corporation (MDRC) to conduct a randomized evaluation comparing the JF program with the earlier state AFDC program for low-income single parents with children. Table 1 provides a detailed summary of the JF and AFDC program features, which we now describe in detail.

A. *Changes in the Treatment of Earnings*

A primary feature of the JF reform was the enhancement of financial incentives to work while on assistance. In the determination of welfare eligibility and grant amounts, Connecticut AFDC recipients faced a fixed earnings disregard of $120 per

Table 1—Summary of Differences between Status Quo and Jobs First Policy Regimes

	Jobs First	Status Quo
Welfare		
Name of program	Temporary Family Assistance (JF-TFA)	Aid for Families with Dependent Children (AFDC)
Eligibility	Earnings below poverty line	Earnings below the level at which benefits are exhausted (see disregard parameters below)
Earnings disregards		
Fixed	N/A	\$120/mo. (first 12 months of work), \$90/mo. (afterward)
Proportional	100%	51% (first 4 months of work), 27% (afterward)
Time limit	21 months	None
Work requirements	Mandatory work first employment services (exempt if child <1)	Education/training (exempt if child <2)
Other features		
Sanctions	Grant reduced by 20%, 35%, or 100% depending on the infraction (first, second, or third) for three months; moderately enforced	Grant reduced to reflect removal of adult from assistance unit; rarely enforced
Asset limit	\$3,000	\$1,000
Family cap	\$50/mo.	\$100/mo.
Transitional Medicaid	2 years	1 year
Transitional child care	Indefinitely as long as income is <75% of state median	1 year as long as income is <75% of state median
Child support	\$100/mo. disregarded; full pass-through	\$50/mo. disregarded; \$50/mo. maximum pass-through
Food Stamps (*if joint with welfare*)		
Earnings disregards		
Fixed	N/A	N/A
Proportional	100% of earnings up to FPL	76% of earnings up to the eligibility threshold

Notes: The Jobs First column pertains to the experimental policy regime while the Status Quo column pertains to the control policy regime. The Welfare panel reports on the rules of the regime-specific welfare policy while the Food Stamps panel reports on the rules of the regime-specific Foods Stamps (FS) program when taken up jointly with welfare assistance (as a stand-alone program Foods Stamps is the same under both policy regimes). Specifically, Welfare is the "family group" variant of welfare assistance considered in the JF experiment called, respectively, AFDC under the status quo and JF-TFA under Jobs First. Categorical eligibility for both AFDC and JF-TFA requires the presence of children below the age of 18 (19 if enrolled in high school) or of a pregnant woman. Connecticut's implementation of AFDC reflects the "fill-the-gap" provision whereby the effective implicit tax rate on earnings is always less than 100 percent. Specifically, "fill-the-gap" budgeting lowers the implicit AFDC tax rate on earnings by a factor of 0.73. For example, in the first four months of employment while on AFDC the usual tax rate would be two-thirds (as part of the so called "$\$30 + 1/3$ policy") but in Connecticut it is $0.73 \times (1 - 1/3) \times 100 = 49$ percent, hence a proportional disregard of 51 percent ensues. From the fifth month forward the usual tax rate would be 100 percent but in CT it is $0.73 \times (1 - 0) \times 100 = 73$ percent, hence a proportional disregard of 27 percent ensues. Certain families, such as those in which the parent is incapacitated, are exempt from the Job First time limits. JF-TFA recipients who reach the 21-month time limit may receive renewable 6-month extensions of their benefits if they have made a good-faith effort to find employment. Both AFDC and JF-TFA impose work requirements. Unless they are exempt, JF-TFA recipients are required to look for a job, either on their own or through Job Search Skills Training courses that teach job-seeking skills. Education and training are generally restricted to those who are unable to find a job despite lengthy up-front job search activities. Job Connection, the state's Job Opportunities and Basic Skills Training program under AFDC, served a small proportion of the total welfare caseload in any month, and a large proportion of those who participated were in education and training activities. Under both AFDC and JF-TFA, sanctions are imposed for failure to comply with employment-related mandates (work requirements). Sanctions entail grant reductions that become more severe as the recipient accumulates instances of noncompliance. During the study period, a JF-TFA recipient's cash grant was reduced by 20 percent for three months in response to the first instance of noncompliance and by 35 percent for three months in response to the second instance. A third instance resulted in cancellation of the entire grant for three months. Under AFDC, a sanction removed the noncompliant individual from the grant. "Full pass-through" of child support means that under JF-TFA all child support collected on behalf of children receiving assistance is given directly to the custodial parent. Under AFDC, when child support was collected, the welfare recipient received a check for the first \$50 that was collected each month (or less than \$50 if less was collected), in addition to her regular welfare check. Family cap is a cap on the benefit increase for children conceived while the mother receives welfare. Transitional Medicaid assistance is Medicaid assistance for families leaving welfare for work. Transitional child care assistance is child care assistance (subsidies) for families leaving welfare for work. AFDC and JF-TFA differ in the duration of the assistance, not in its nature or generosity. Because cash assistance recipients are categorically eligible for FS, the asset and earned income rules effectively apply to FS eligibility while a family receives welfare. In particular, Job First's enhanced earning disregard applies to the FS grant calculation so that all earnings below the poverty line are disregarded. When JF-TFA recipients lose their welfare grant, they also lose the enhanced FS earned income disregard (Bloom et al. 2002).

month for the 12 months following the first month of employment while on assistance and \$90 per month afterward (henceforth, the *unreduced* and *reduced* fixed disregards). They also faced a proportional disregard of any additional earnings: 51 percent for the four months following the first month of employment while on assistance and 27 percent afterward (henceforth, the *unreduced* and *reduced* proportional disregards). By contrast, the JF program disregarded 100 percent of earnings up to the monthly federal poverty line (FPL). This change was meant to incentivize work but also generated an eligibility "notch" in the transfer scheme, with a windfall loss of the entire grant amount occurring if a woman earned \$1 more than the poverty line. This created strong incentives for some women to reduce their earnings in order to maintain eligibility for assistance.

We can formalize the rules governing welfare transfers by means of the transfer function $G_i^t(E)$ which gives the monthly grant amount associated with welfare participation at earnings level E under policy regime $t \in \{a, j\}$ (AFDC or JF, respectively). The i subscript acknowledges that the grant amount also varies according to a woman's history of employment while on assistance (under AFDC only) and the size of a woman's *assistance unit*, which consists of the woman receiving welfare plus any eligible dependent children.[4] Letting $\mathbf{1}[\cdot]$ be an indicator for the expression in brackets being true, the regime-specific transfer functions can be written

$$G_i^a(E) = \mathbf{1}[E \leq \overline{E}_i]\left(\overline{G}_i - \mathbf{1}[E > \delta_i](E - \delta_i)\tau_i\right) \tag{1}$$

$$G_i^j(E) = \mathbf{1}[E \leq FPL_i]\,\overline{G}_i, \tag{2}$$

where $\delta_i \in \{90, 120\}$ and $\tau_i \in \{0.73, 0.49\}$ are the fixed and proportional AFDC earnings disregards, $\overline{G}_i$ is the base grant amount (which is common to JF and AFDC), FPL_i is the federal poverty line, and $\overline{E}_i = \overline{G}_i/\tau_i + \delta_i$ is the so-called AFDC *break-even* earnings level above which a woman becomes ineligible for cash assistance. The i subscripts on $\overline{G}_i$ and FPL_i reflect that these quantities vary with the size of the assistance unit.

Figure 1 provides a stylized depiction of the changes to the cash assistance component of welfare experienced by a woman with two children who, in 1997, would face a base grant amount of \$543 and a federal poverty line of \$1,111. This woman's welfare history gives her access only to the reduced fixed and proportional disregards under AFDC. The figure plots the woman's disposable income (earned income plus welfare assistance) against her earnings E under AFDC (panel A) and JF (panel B), respectively. Transfers under JF exhibit a large discontinuity at the federal poverty line. By contrast, assistance phases out smoothly under AFDC. For instance, if the woman were to earn \$200 per month, she would receive a welfare transfer that is \$80 (17 percent) larger under JF than under AFDC. While if she earned \$800 per month, she would receive a transfer that is \$518 (2,098 percent) larger under JF than under AFDC. Although in Figure 1 the AFDC transfer is fully exhausted at an earnings level $\overline{E}$ that is strictly below the poverty line, this is not always the case. A

[4] Children are eligible if they are under 18 years old or under 19 years old and in school.

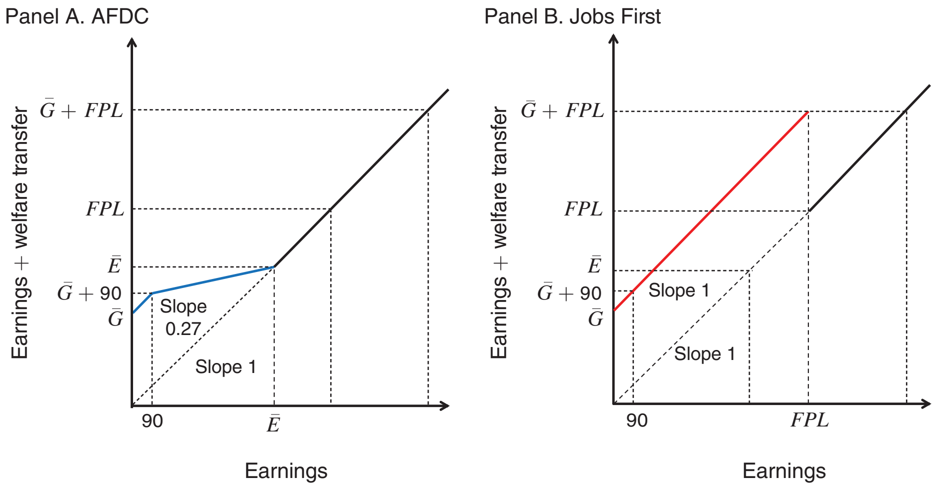


Figure 1. Monthly Earnings Plus Welfare Transfer under AFDC and Jobs First

Notes: The figures (not drawn to scale) depict the sum of monthly earnings and welfare transfers for a woman with 2 children under AFDC (panel A) and Jobs First (panel B) policy rules as of 1997. FPL refers to federal poverty line (\$1,111) and $\overline{G}$ is the base grant amount (\$543). The illustration assumes that the woman only has access to the fixed \$90 disregard and the proportional 73 percent disregard under AFDC which implies that the AFDC transfer is exhausted at earnings level $\overline{E}$ corresponding to $\overline{G}/0.73 + 90$ (\$834). The JF welfare transfer falls to zero at earnings levels above FPL. Under JF, a woman who earns between FPL (\$1,111) and $FPL + \overline{G}$ (\$1,654) can increase the sum of earnings and welfare transfers by taking up welfare assistance and working less.

woman with access to the unreduced proportional and fixed disregards exhausts her AFDC transfer at an earnings level slightly above the poverty line.

Welfare is, of course, part of a broader web of tax and transfer programs. Figure 2 depicts the woman's monthly income accounting for the Food Stamps (FS) program, payroll and Medicaid taxes, and the Earned Income Tax Credit (EITC). The Food Stamps program interacts with welfare assistance both because welfare recipients are categorically eligible for the program and because welfare transfers are treated as income in the determination of the food stamps transfer amounts. The JF reform introduced a further link between cash and in-kind assistance: conditional on joint take-up, earnings up to the poverty line were disregarded in the determination of both the welfare and the Food Stamps transfers. This feature is clearly visible in Figure 2—under JF, the combined welfare and FS transfer depends only on whether earnings exceed the poverty line, in which case assistance is denied. Thus, JF's impact on the Food Stamps program amplifies the notch at the poverty line.[5]

B. *Work Requirements, Sanctions, and Time Limits*

At the time of the reform, Connecticut mandated work requirements for all AFDC recipients except those with a child under age two (who were exempt). AFDC work requirements could be met by paid employment or, in place of employment, by participation in employment-related services. The MDRC final report describes these

[5] The EITC and other taxes do not directly interact with cash and in-kind assistance because income from welfare and food stamps are not counted in the determination of taxes and tax credits.

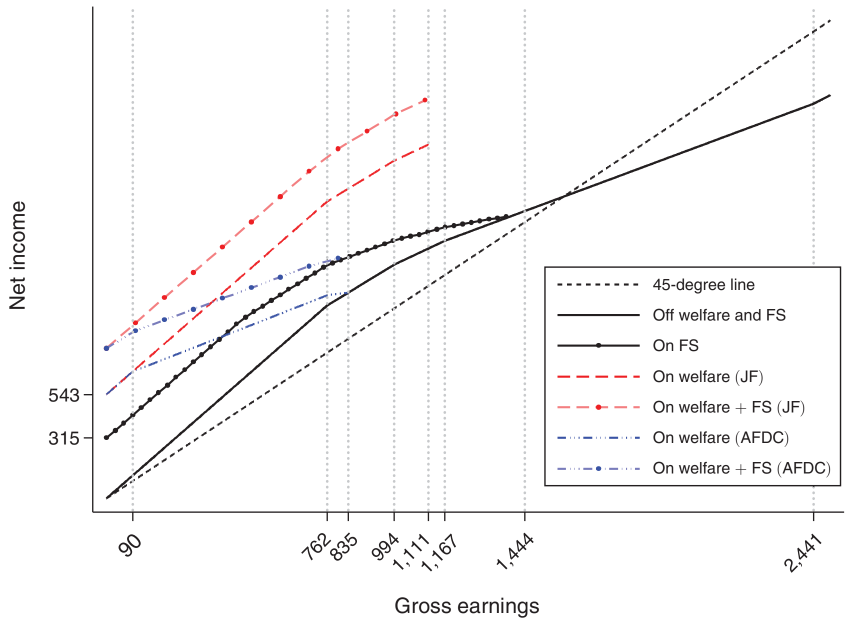


Figure 2. Net Income under Status Quo and Jobs First Policies, Accounting for Food Stamps and Taxes

Notes: Figure depicts net income as a function of gross earnings for an assistance unit of size 3 in 1997 under the status quo policy and the JF policy rules. Illustration assumes household only has access to fixed \$90 disregard under AFDC and faces \$366 in monthly rental expenses. Net income is earnings net of federal income taxes and inclusive of EITC and welfare and food stamps transfers (given participation in either program). Vertical lines: at the AFDC fixed earning disregard and break-even level (\$90 and \$835); at the end of the EITC's phase-in and start and end of the phase-out regions (\$762, \$994 and \$2,441); at the minimum taxable earnings (\$1,167); at the FPL (\$1,111); and at 1.3 × FPL (\$1,444), which is a FS eligibility threshold under AFDC. Horizontal ticks: at maximum FS and welfare grants.

services as "a small-scale, largely voluntary, education-focused welfare-to-work program" (Bloom et al. 2002, p. 28) with lax enforcement. JF recipients, by contrast, were required to participate in employment services targeted toward quick job placement unless they were parents caring for a child less than one year old.[6] Additionally, the JF reform stepped up sanctions for noncompliance with work requirements. JF recipients who failed to make good faith efforts to find work while receiving assistance could be sanctioned by having their welfare grant reduced or temporarily canceled. Under AFDC, sanctions involved removing the noncompliant adult from the grant calculation rather than closing the entire case.

[6] Bloom et al. (2002, p. 31) note that "Connecticut, like many other states, did not strongly enforce the existing requirements for AFDC recipients to participate in employment-related activities (in fact there were waiting lists for services). Job Connection, the states Job Opportunities and Basic Skills Training (JOBS) program, served a small proportion of the total welfare caseload in any month, and a large proportion of those who participated were in education and training activities." By contrast, the JF work mandates appear to have been implemented strictly with minimal emphasis on training: "Nearly all [non-exempted] JF participants were required to begin by looking for a job, either on their own or through Job Search Skills Training (JSST), a group activity that teaches job-seeking and job-holding skills. Education and training were generally reserved for recipients who were unable to find a job despite lengthy up-front job search activities" (Bloom et al. 2002, p. 11).

Finally, women could remain on welfare indefinitely under AFDC provided that their children were of eligible age. By contrast, under JF women were limited to a lifetime total of 21 months of cash assistance. However, months when women were exempt from work requirements did not count toward the lifetime limit and additional exemptions from the time limit were available in some cases (e.g., if the parent was incapacitated or caring for a disabled child). Moreover, six-month extensions from the time limit were possible if recipients were deemed to have made a "good faith" effort to find employment. Bloom et al. (2002, p. 37) report that "in general, a good-faith effort was assumed as long as the recipient was not sanctioned more than once and did not quit a job without 'good cause' in the final six months of assistance." There was no limit on the number of six-month extensions a family could receive. Survey evidence from Bloom et al. (2002) suggests that, in practice, a majority of the cases reaching the time limit were granted an extension and, during the first year after random assignment, nearly 20 percent of the JF units were exempt from time limits.

C. *Other Changes*

The JF reform also entailed some minor changes to programs available to women leaving welfare. Under AFDC, recipients were eligible for 12 months of Transitional Child Care subsidies if they left welfare for work, while under JF, cases were eligible for child care subsidies indefinitely provided that their income did not exceed 75 percent of the state median income. Likewise, under AFDC, assistance units leaving welfare because of increased earnings were eligible for one year of Transitional Medicaid, while under JF, units were eligible for two years of Medicaid. While these programs could create additional incentives to work, Bloom et al. (2002) argue that these components of the JF reform had little impact on actual access to child care or health care because of contemporaneous state-level programs covering essentially the same population.[7]

JF also changed the treatment of income received in the form of child support transfers. Under AFDC, recipients received only the first $50 of the child support collected by the child support collection agency from the child's noncustodial father. The amount received was then disregarded in the computation of the welfare grant. Under JF, recipients received a check for the full amount of any child support collected with only the first $100 disregarded in computing the welfare transfer. These changes may have induced income effects since women whose child support collection was between $50 and $100 could enjoy an increased welfare transfer with no change in behavior. However, these income effects are likely negligible given that

[7] Regarding Transitional Child Care (TCC), Bloom et al. (2002, p.161) write that, "in practice, however, the difference between these two policies was minimal, because AFDC members who reached the end of their eligibility for TCC could move directly into the child care certificate program (that is, income-eligible child care) for low-income working parents." As to the effects of Transitional Medicaid, they write that "the magnitude of the treatment difference related to medical assistance has diminished over time, as Connecticut has expanded the availability of health coverage to low-income children and adults who do not receive welfare" (p. 133). In addition, they note that "the 1996 federal welfare law 'de-linked' eligibility for Medicaid from eligibility for welfare and created a new coverage category for families who are not on welfare but who meet the AFDC eligibility criteria that were in place in July 1996. These statewide expansions in health coverage for children and adults are available to both the JF group and the AFDC group" (p. 47). Taken together, these observations suggests that the additional 12 months of Transitional Medicaid available under JF are unlikely to have induced changes in the value of working off assistance.

they only apply to women within this restricted range of child support payments (payments above $100 were deducted dollar for dollar from benefits) and since the amount of additional income per month was very small.

II. Data and Descriptive Statistics

Before delving into a model of a woman's response to the JF reform, it is useful to introduce some basic features of the data. Between January 1996 and February 1997, MDRC randomly assigned 4,803 welfare recipients and applicants to either an AFDC "control" group or to a JF "experimental" group. Data were collected on these women through the end of 2000 when the experiment ended. We next describe the MDRC Jobs First Public Use Files used in our analysis and briefly examine the baseline characteristics of our estimation sample. We then examine the distribution of earned income relative to the eligibility notch in the experimental sample.

A. *Data*

The MDRC Jobs First public use files contain a baseline survey of demographic and family composition variables merged with longitudinal administrative information on welfare and food stamps participation, rounded welfare and food stamps payments, and rounded earnings covered by the state unemployment insurance (UI) system. There are a number of limitations to these data. While participation and transfers are measured monthly, UI earnings data are available only quarterly. Data on hours and weeks worked are not available, which prevents us from inferring hourly wages. Also, earnings reported to the welfare agency by applicants and recipients are not available.

Another difficulty is that the administrative measure of assistance unit size is missing for most cases. This is problematic because a woman's assistance unit size determines her poverty line and the corresponding location of the JF eligibility notch she faces. In the JF sample, we are able to infer an assistance unit size from the grant amount in months when a woman is on welfare. But in the AFDC sample, the grant amount depends on the woman's history of past employment and welfare take-up, which we observe only partially. Consequently, we cannot reliably infer an assistance unit size from grant amounts under AFDC. For this reason, when computing treatment effects by assistance unit size, we rely on a variable collected in the baseline survey named "kidcount" that records the number of children in the household at the time of random assignment. As might be expected, using the kidcount variable leads to underestimates of true assistance unit size since women may have additional children over the seven quarters following the baseline survey. To deal with this problem we inflate the kidcount based measure of assistance unit size by one in order to avoid understating the location of the poverty line for most women.[8] Additional details regarding variable construction are provided in the online Appendix.

[8]Understating the poverty line could lead to an overestimate of the population at risk of opting in to welfare. We err on the side of overstating the poverty line because one of our goals is to provide a conservative assessment of whether opt-in behavior actually takes place. Online Appendix Table A1 tabulates the kidcount variable against

B. *Baseline Characteristics of the Analysis Sample*

Table 2 provides baseline descriptive statistics for our analysis sample. We have 4,642 cases with complete prerandom assignment data and nonmissing values of the kidcount variable. There are some mildly significant differences between the AFDC and JF groups in their baseline characteristics, however these differences are not jointly significant. We follow BGH in using propensity score reweighting to adjust for these baseline differences.[9] After adjustment, the baseline means of the AFDC and JF groups are very similar. We also examine two subgroups defined by whether they had positive earnings seven quarters prior to random assignment (the two right-most panels in Table 2). Because preassignment earnings proxy for tastes and earnings ability, the JF reform likely presented these groups with different incentives, which makes them useful for exploring treatment effect heterogeneity (see BGH 2014 for a related subgroup analysis).

C. *Earnings Distribution*

Panels A through C of Figure 3 provide histograms by assistance status of earned income in the JF sample for the seven quarters following random assignment—a horizon over which no case was in danger of reaching the time limit. Earnings are rescaled relative to three times the federal poverty line, which is the maximum amount that a woman can earn in one quarter while maintaining welfare eligibility throughout the quarter. Each woman's poverty line is determined using our administratively inferred measure of assistance unit size.

Many labor supply models predict bunching of earnings at notches (Slemrod 2010; Kleven and Waseem 2013). Like BGH (2006), we find no evidence of such bunching at the JF eligibility notch (panel A of Figure 3). Rather, the earnings density declines smoothly through the notch which should bound, to its right, a dominated earnings region. Compared to women not on welfare in the quarter (panel C), there is arguably an excess "mound" in the density of earnings below the notch for women on welfare throughout the quarter (panel B). While it is possible to rationalize the absence of bunching with certain distributions of preferences, this evidence is also consistent with the possibility that women face significant labor supply constraints—a conjecture that has received substantial empirical support in related settings (Altonji and Paxson 1992; Dickens and Lundberg 1993; Chetty, Friedman, Olsen, and Pistaferri 2011; Beffy et al. 2014).

A conspicuous feature of panel B is that the distribution of earnings stretches well beyond the poverty line, despite the fact that women with such earnings levels should be ineligible for welfare under JF. While it is possible that some of these observations are the result of measurement problems, underreporting behavior is

the administrative measure available in the JF sample. Our inflation scheme maps the kidcount measure to its modal administrative value plus one. As detailed in online Appendix Table A7, our main results are robust to alternate codings such as inflating the assistance unit size by two and not inflating it at all.

[9] These techniques are described in the online Appendix. The baseline sample in BGH (2006) contains 4,803 cases. Relative to their analysis, we impose the additional restriction that the kidcount variable be nonmissing. We also drop one AFDC case from our analysis with unrealistically high quarterly earnings that sometimes led to erratic results.

Table 2—Mean Sample Characteristics

	Overall sample				Zero earnings Q7 pre-RA		Positive earnings Q7 pre-RA	
	Jobs First	AFDC	Difference	Difference (adjusted)	Jobs First	AFDC	Jobs First	AFDC
Demographic characteristics								
White	0.374	0.360	0.014	0.001	0.340	0.331	0.453	0.421
Black	0.380	0.384	−0.004	0.000	0.370	0.360	0.404	0.435
Hispanic	0.214	0.224	−0.010	−0.001	0.258	0.275	0.110	0.117
Never married	0.654	0.661	−0.007	0.000	0.658	0.654	0.645	0.674
Div./wid./sep./living apart	0.332	0.327	0.005	0.000	0.327	0.334	0.345	0.312
High school dropout	0.350	0.334	0.017	0.000	0.390	0.394	0.257	0.209
High school diploma/GED	0.583	0.604	−0.021	0.000	0.550	0.555	0.661	0.706
More than high school diploma	0.066	0.062	0.004	0.000	0.060	0.051	0.082	0.085
More than two children	0.235	0.214	0.021	0.000	0.260	0.250	0.176	0.139
Mother less than 25 years old	0.287	0.298	−0.011	−0.003	0.287	0.268	0.288	0.361
Mother 25–34 years old	0.412	0.414	−0.003	0.005	0.410	0.419	0.416	0.405
Mother more than 34 years old	0.301	0.287	0.014	−0.002	0.303	0.313	0.297	0.233
Average quarterly pretreatment values								
Earnings	673	750	−76*	4	174	185	1,856	1,935
	[1,306]	[1,379]	(40)	(6)	[465]	[479]	[1,802]	[1,828]
Cash welfare	903	845	58**	−1	1,050	1,022	555	475
	[805]	[784]	(23)	(2)	[811]	[799]	[677]	[602]
Food stamps	356	344	12	0	399	398	253	230
	[320]	[304]	(9)	(1)	[326]	[310]	[282]	[256]
Fraction of pretreatment quarters with								
Any earnings	0.319	0.347	−0.029***	0.000	0.137	0.143	0.751	0.776
	[0.361]	[0.370]	(0.011)	(0.001)	[0.211]	[0.215]	[0.262]	[0.239]
Any welfare assistance	0.581	0.551	0.030*	−0.001	0.650	0.636	0.418	0.373
	[0.451]	[0.449]	(0.013)	(0.001)	[0.439]	[0.439]	[0.438]	[0.416]
Any food stamp assistance	0.613	0.605	0.008	0.000	0.670	0.674	0.480	0.460
	[0.437]	[0.431]	(0.012)	(0.001)	[0.427]	[0.421]	[0.433]	[0.418]
Cases	2,318	2,324			1,630	1,574	688	750

Notes: Sample units missing baseline data on number of children (kidcount) are excluded. The included units are organized into the following subsamples: Overall sample/Jobs First refers to units in the experimental sample; Overall sample/AFDC refers to units in the control sample; Zero earnings Q7 pre-RA/Jobs First refers to units in the experimental sample that have zero earnings in the seventh quarter prior to random assignment (RA); Zero earnings Q7 pre-RA/AFDC refers to units in the control sample that have zero earnings in the seventh quarter prior to RA; Positive earnings Q7 pre-RA/Jobs First refers to units in the experimental sample that have positive earnings in the seventh quarter prior to RA; and Positive earnings Q7 pre-RA/AFDC refers to units in the control sample that have positive earnings in the seventh quarter prior to RA. Adjusted differences are computed via propensity score reweighting. Numbers in brackets are standard deviations and numbers in parentheses are standard errors calculated via 1,000 block bootstrap replications (resampling at case level). Significance indicators provided only for difference estimates.

***Significant at the 1 percent level.
**Significant at the 5 percent level.
*Significant at the 10 percent level.

also undoubtedly at play here. The MDRC final report (Bloom et al. 2002, p. 38) provides some direct evidence on this point, noting that, in the AFDC group, the fraction of women with earnings in the UI system was about ten percentage points higher than the fraction reporting earnings to the welfare agency. In the JF group, the fraction reporting earnings to the welfare system was nearly identical to the fraction with UI earnings. However, this may be an artifact of the 100 percent JF earnings disregard which creates incentives to report an earnings amount below the poverty line rather than no earnings at all. Evidence on such partial underreporting

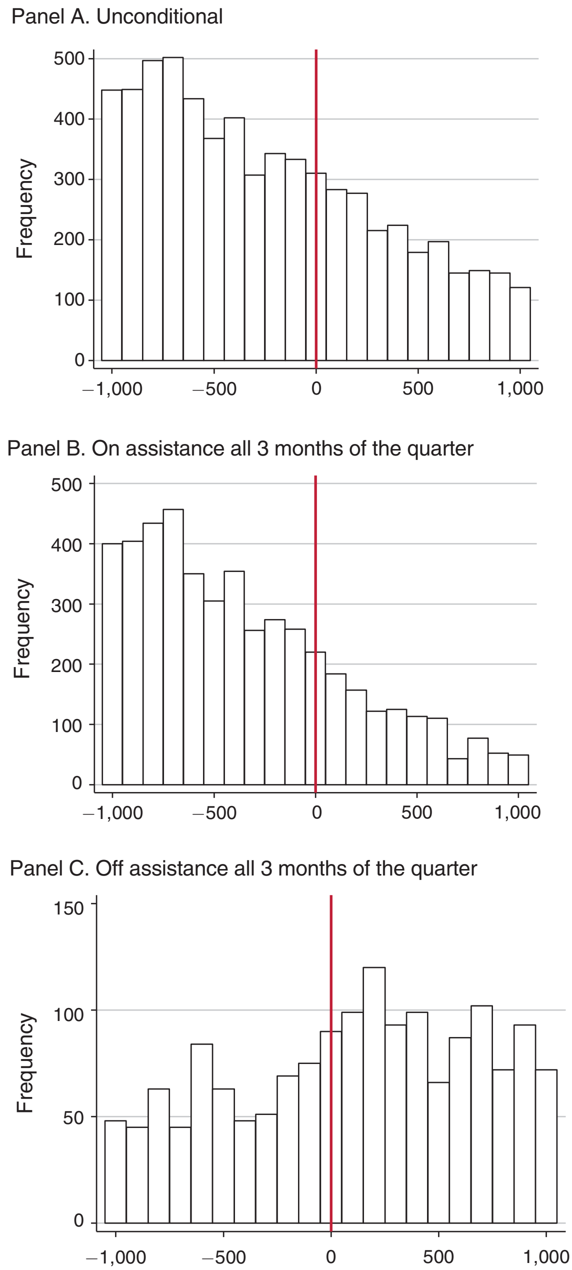


Figure 3. Distribution of Quarterly Earnings Centered at Three Times Monthly Federal Poverty Line

Notes: Restricted to the Jobs First sample in quarters 1–7 post random assignment. Assistance unit size has been inferred from monthly transfer payment (sizes above eight have been excluded). The bins in the histograms are \$100 wide with bin 0 containing three times the monthly federal poverty line corresponding to the assistance unit's size and the calendar year of the quarterly observation. Vertical line indicates Jobs First eligibility threshold at three times the monthly federal poverty line.

was found in a related context by Hotz, Mullin, and Scholz (2003), who analyzed data from a welfare reform experiment in California.[10]

III. Earnings Impacts and a Test for Intensive Margin Responsiveness

As previously documented by BGH, the JF reform had nuanced effects on the distribution of earned income. Here we briefly review those impacts and then ask whether they could have been generated solely by extensive margin responses.

A. *Earnings Impacts*

Panel A of Figure 4 provides reweighted empirical distribution functions (EDF) by experimental status of rescaled earnings in the first seven quarters of the experiment. The poverty line employed in the rescaling is computed using our survey-based measure of assistance unit size. Significant opt-in behavior should lead earnings levels below the poverty line to be more common in the JF sample than the AFDC sample.

A reweighted Kolmogorov-Smirnov test strongly rejects the null hypothesis that the two EDFs are identical. More quarters exhibit positive earnings in the JF sample than in the AFDC sample, indicating that JF successfully incentivized many women to work.[11] The earnings EDF rises more quickly in the JF sample than in the AFDC sample, signaling excess mass at low earnings levels. Also, the EDFs cross below the notch, implying that the fraction earning less than the poverty line is slightly greater in the JF sample than among the AFDC controls. A large increase in the fraction earning less than the poverty line would be suggestive evidence of an opt-in response, however the impact here is small and statistically insignificant.

Panels B and C of Figure 4 provide corresponding EDFs in the two subsamples defined by their earnings in the seventh quarter prior to random assignment. These groups are of interest because prerandom assignment earnings are a strong predictor of postrandom assignment earnings and therefore proxy for the relevant range of the budget set an agent would face under AFDC. Accordingly, units with positive prerandom assignment earnings should be most likely to exhibit an opt-in effect, while units with zero earning should be more likely to be pushed into the labor force by JF. The figures confirm that the expected pattern of heterogeneity is in fact present: the positive earnings group experienced less of an impact on the fraction of quarters spent working and a significant increase in the fraction of quarters with earnings less than or equal to the poverty line. The zero earnings group, by contrast, exhibits a large increase in the fraction of quarters working, but essentially no impact on the fraction of quarters with earnings less than or equal to the poverty line.

[10] Comparing administrative earnings records from the California unemployment insurance system with earnings reported to welfare, they find that about one-quarter of welfare cases report earning amounts to the welfare agency that are lower than the figures recorded in the state UI system. Among these cases, the average fraction of UI earnings reported varied from 64 percent to 84 percent depending on the year studied.

[11] Online Appendix Table A2 provides standard errors on selected earnings impacts, which confirm the visual impression of panels A–C of Figure 4.

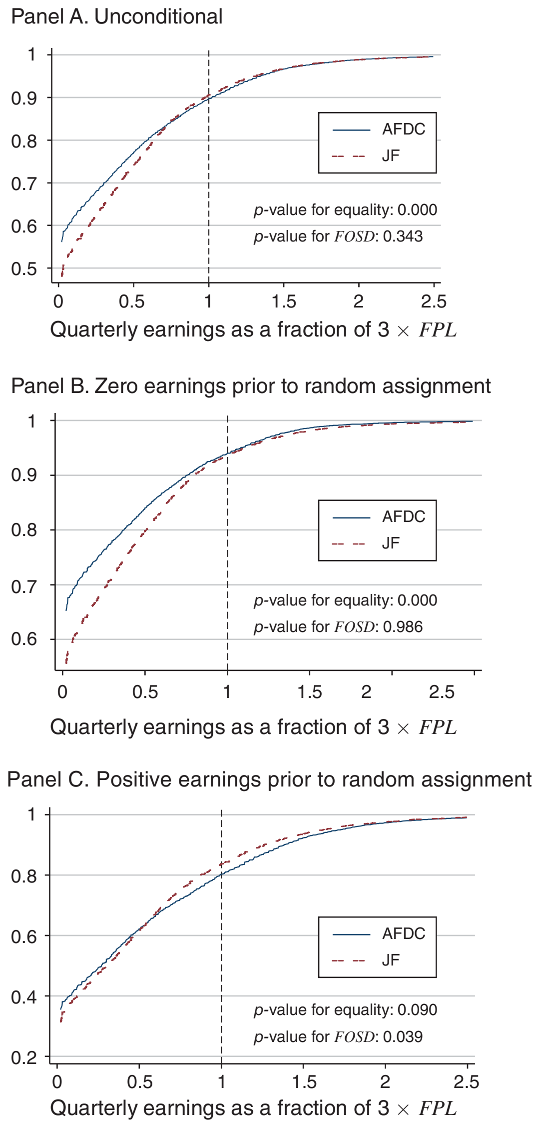


Figure 4. EDFs of Quarterly Earnings Relative to Three Times Federal Poverty Line

Notes: Figures give reweighted empirical distribution functions of quarterly UI earnings (in quarters 1–7 postrandomization) in JF and AFDC samples relative to three times the monthly federal poverty line. Assistance unit size determined by baseline survey variable "kidcount" (see text for details). Panel B refers to women with zero earnings in the seventh quarter prior to random assignment, while panel C refers to women with positive earnings in that quarter. "*p*-value for equality" refers to a Kolmogorov-Smirnov test of quality of the two distributions, while "*p*-value for FOSD" refers to a Barrett-Donald test for first-order stochastic dominance (FOSD) of the JF distribution over the AFDC distribution (both based on 1,000 block bootstrap replications at case level: see online Appendix for details).

B. *A Test for Intensive Margin Responsiveness*

Could these distributional impacts on earnings have been generated by extensive margin adjustments alone? In the absence of intensive margin adjustments, the reform simply shifts women from not working at all to earning positive amounts, which implies the distribution of earned income in the JF sample should stochastically dominate the distribution in the AFDC sample.[12] Using a variant of the testing procedure of Barrett and Donald (2003) described in the online Appendix, we fail to reject the null hypothesis that the JF earnings distribution stochastically dominates the earnings distribution in the pooled AFDC sample (panel A of Figure 4). However, first-order stochastic dominance is rejected at the 5 percent level in the positive earnings sample (panel C), indicating that intensive margin responses did in fact occur in response to the reform, but are difficult to detect in the pooled sample using earnings alone.

IV. Model

Having established the presence of both intensive and extensive margin labor supply responses to the JF reform, we now seek to infer the frequency of these responses. What fraction of women were induced to lower their earnings and take up welfare in response to the JF reform? What share of women were induced to work at earnings levels above the poverty line? How many women were induced to leave welfare? The fundamental challenge to answering such questions is that we cannot observe the choice each woman would have made under the policy regime to which she was not assigned.

In this section we develop an optimizing model that formalizes the incentives provided by the JF reform and restricts the set of possible labor supply and program participation responses to the experiment. We depart from conventional structural modeling approaches (e.g., Moffitt 1983; Keane and Moffitt 1998; Hoynes 1996; Swann 2005; Keane and Wolpin 2002, 2007, 2010; Chan 2013) by allowing for a nonparametric specification of preferences. Motivated by our finding of the absence of a spike in the earnings distribution at the JF eligibility threshold, we allow for the possibility that women face constraints on their labor supply decisions. Women also choose how much of their earnings to report to the welfare agency, creating the possibility that earnings-ineligible women are welfare participants.

Our analysis relies on a number of simplifying assumptions. First, the model is static. In practice, women are likely to make choices taking into account both current and future payoffs. For our purposes, these motives are only problematic if they rationalize responses that do not emerge under myopic decision making. For this to be the case, alternative specific continuation values would need to differ across AFDC and JF in ways that undermine our static conclusions regarding which choices are made more or less attractive by the reform. The JF time limits are the most obvious culprit for such effects since they could make working while on welfare less attractive under JF than under AFDC. However, as described

[12] First-order stochastic dominance implies the absence of negative QTEs. Therefore, the analysis of BGH already provides evidence against the extensive margin-only null hypothesis. However, focusing on particular QTEs that happen to be significant can generate a multiple testing problem. The methods used here address this problem.

in Section X, an adaptation of the Grogger and Michalopoulos (2003) test for anticipatory responses fails to detect forward-looking behavior, leading us to believe that the dynamic incentives of the reform are in fact weak in this sample.[13] Second, the model ignores the Transitional Child Care, Transitional Medicare, and child support components of the JF reform. We explained above why these features of the reform likely had minimal effects. Introducing them would substantially complicate our analysis and add little given that we lack data on participation in these programs. Third, to simplify exposition, the model ignores the Food Stamps program, payroll and Medicare taxes, and the EITC. We explain in Section VI why extending the model to incorporate these policies has no effect on our revealed preference arguments. Finally, we present the model for a woman whose break-even earnings level $(\overline{E}_i)$ under AFDC is below the federal poverty line as in Figure 1. In Section VI, we discuss how things differ for women who have access to unreduced earnings disregards yielding $\overline{E}_i > FPL_i$.

A. *The Decision Problem*

Consider a woman with children, call her i, subject to a policy regime indexed by $t \in \{a, j\}$ (AFDC or JF, respectively). In a given month, woman i samples $K_i \geq 0$ job offers, composed of wage and hours offer pairs: $\Theta_i \equiv \left\{\left(W_i^k, H_i^k\right)\right\}_{k=1}^{K_i}$. The woman's offer set Θ_i reflects a mix of luck and the woman's labor market skills. Woman i decides which (if any) of the K_i offers to accept, whether to participate in welfare (represented by the indicator $D \in \{0, 1\}$), and a level $\left(E^r \geq 0\right)$ of earned income to report to the welfare agency. We assume E^r is less than or equal to her actual earnings $E = WH$ where W and H refer to the wage and hours at her chosen job (which are both zero when no offer is accepted).[14]

Woman i consumes her earnings plus any welfare transfer. Specifically, her consumption is given by

$$C = C_i^t(E, D, E^r) = E + DG_i^t(E^r), \tag{3}$$

where the welfare grant $G_i^t(E^r)$ is determined according to the regime-specific transfer functions (1)–(2) based upon her reported (as opposed to her actual) earnings.

Woman i's preferences are represented by the utility function

$$U_i^t(H, C, D, Z, R), \tag{4}$$

[13] Returns to labor market experience are a second culprit. Our model posits regime-invariant earning offer functions, which implies that the attractiveness of off-welfare alternatives is assumed to be the same under AFDC and JF. If JF induces more women to work, and if returns to labor market experience are substantial, this assumption is violated. However, the magnitude of experience effects in our sample is likely to be small. For example, after studying data from a similar welfare experiment—the Canadian Self Sufficiency Project (SSP)—Card and Hyslop (2005, p. 1740) conclude that "work experience attributable to SSP appears to have had no detectable effect on wage opportunities." Couch (2014) uses 14 years of postrandomization earnings data from the JF reform and concludes that "the short-term intervention did not appear to have altered the long-term outcomes of participants examined in terms of employment or labor market earnings."

[14] Allowing overreporting behavior would essentially nullify the JF work requirements. In practice, concocting a fictitious job was difficult as employment had to be verified by case workers.

where $Z = Z(D, E^r) = D\mathbf{1}[E^r = 0]$ is an indicator that equals 1 if she reports zero earnings to the welfare agency and $R = R(E, D, E^r) = D\mathbf{1}[E^r < E]$ is an indicator that equals 1 if she underreports her earnings to the welfare agency. The dependence of utility on D captures the potential for a "stigma" (or, conversely, a psychic benefit) associated with welfare participation (Moffitt 1983), while the dependence on Z captures the "hassle" associated with welfare work requirements. The dependence of utility on R captures the cost of underreporting to the welfare agency which may reflect the effort exerted in disguising earnings and the possibility of being caught underreporting.[15] Utility is indexed by the policy regime t to allow for differences in hassle disutility under AFDC and JF.

We assume that the utility function in (4) obeys the following restrictions:

$$\text{(A1)} \qquad \text{utility is strictly increasing in } C;$$

$$\text{(A2)} \qquad U_i^t(H, C, 1, Z, 1) < U_i^t(H, C, 1, Z, 0) \quad \text{for all } t \in \{a, j\};$$

$$\text{(A3)} \qquad U_i^t(H, C, 1, 1, R) \leq U_i^t(H, C, 1, 0, R) \quad \text{for all } t \in \{a, j\};$$

$$\text{(A4)} \qquad U_i^j(H, C, 1, 1, R) \leq U_i^a(H, C, 1, 1, R);$$

$$\text{(A5)} \qquad U_i^j(H, C, 1, 0, R) = U_i^a(H, C, 1, 0, R);$$

$$\text{(A6)} \qquad U_i^j(H, C, 0, 0, 0) = U_i^a(H, C, 0, 0, 0).$$

Assumption (A1) is a standard nonsatiation condition. Assumption (A2) states that underreporting is costly under either policy regime. Assumption (A3) states that reporting zero earnings weakly lowers utility (due to welfare hassle). Assumptions (A4–A6) formalize our institutional knowledge of the JF reform. In accord with JF's increased work requirements, (A4) restricts the utility of reporting zero earnings while on welfare to be no higher under JF than AFDC. Assumption (A5) restricts both the utility costs of underreporting and the psychic costs (or benefits) of welfare participation to be regime-invariant among recipients who report positive earnings. Finally, assumption (A6) requires utility to be regime-invariant when off assistance.

The specification above of utility is extremely general. Due to the nonseparability of H and C, leisure and consumption may be complements or substitutes and preferences may be nonhomothetic as in classic Stone-Geary specifications of utility. Because we do not require monotonicity with respect to H, woman i may value working full-time more than working part-time or vice versa. Likewise, participation in welfare may increase or decrease her utility. Welfare stigma creates the possibility that woman i refuses assistance despite being eligible. The effect of welfare participation on her utility is allowed to vary with consumption and leisure due to the nonseparability of D. Similarly, the hassle disutility is allowed to vary with

[15] An important restriction here is that the cost of underreporting does not depend on what fraction of earnings are underreported. This feature of our model is necessitated by the fact that reported earnings E^r are not included in MDRC's Public Use Files. See Saez (2010) for a similar restriction involving a fixed "moral" cost of misreporting income to tax authorities.

consumption and leisure due to the further nonseparability of Z. Note that we have not assumed continuity of utility with respect to H or C, which accommodates the possibility that woman i faces a fixed cost (or benefit) of work such as a monthly commuting cost. Fixed costs discourage work at low earnings levels and create the possibility that she responds to marginal changes in work incentives by earning large amounts instead of not working at all (Cogan 1981).

A special case of (4) monetizes welfare stigma, hassle, the disutility of underreporting, and the (dis)utility of working as follows:

$$U_i\big(H, C - \phi_i D - \eta_i^t Z - \kappa_i R - \mu_i \mathbf{1}[E > 0]\big), \tag{5}$$

where ϕ_i is the monetized cost of welfare stigma, η_i^t is the hassle cost of reporting zero earnings under regime t, κ_i is the cost of underreporting, and μ_i is a fixed cost (or benefit) of work. The parameters $\left(\phi_i, \eta_i^a, \eta_i^j, \kappa_i, \mu_i\right)$ inherit the restrictions above on preferences. Specifically, (A2) and (A4) imply $\kappa_i > 0$ is regime-invariant, (A3) and (A4) stipulate that $\eta_i^j \geq \eta_i^a \geq 0$, and (A5) requires that ϕ_i be regime-invariant. Finally, in accordance with (A6), μ_i and the two-argument utility function in (5) are both regime-invariant. We refer to the second argument of (5) as the "consumption equivalent." We selectively consider this "monetized" specification below to aid in illustrating the mechanics of the model and the implications of further structuring preferences. Our main results rely on the more general specification given in (4).

Woman i's objective is to maximize her utility under policy regime t. Hence, she selects a labor supply, program participation, and reporting alternative X_i^{t*} that belongs to[16]

$$\underset{(W,H)\in\{\Theta_i,(0,0)\},\ D\in\{0,1\},\ E^r\in[0,E]}{\arg\max} \quad U_i^t\big(H, C_i^t(WH, D, E^r), D, Z(D, E^r), R\big(WH, D, E^r\big)\big). \tag{6}$$

We refer to X_i^{t*} as woman i's *choice* under policy regime t. Note that her pair $\left(X_i^{a*}, X_i^{j*}\right)$ of regime-dependent choices is governed by the vector of primitives,

$$\theta_i \equiv \left(U_i^j(.,.,.,.,.), U_i^a(.,.,.,.,.), \Theta_i, \bar{G}_i, \delta_i, \tau_i\right).$$

We provide some examples of how choices depend on these primitives at the end of the next section.

B. *Optimal Reporting*

Recall from (4) that woman i's utility depends on whether or not she underreports her earnings to the welfare agency but not on the magnitude of the underreporting. As a result, optimal reporting obeys a particularly convenient decision rule. Specifically, assumptions (A1), (A2), and (A3) imply that when concealing her earnings optimally, woman i reports a positive amount to the welfare agency

[16] Indifferences between alternatives may arise that lead the arg max to be a set instead of a vector. We do not model how woman i chooses among alternatives between which she is indifferent. We only assume that the rule she uses to choose among them is invariant to the policy regime t.

that allows her to avoid any hassle penalty and to receive a transfer $\bar{G}_i$ irrespective of the policy regime she faces (see Lemma 2 in the online Appendix). Hence, by assumption (A5), for any level of actual earnings E, the utility she receives from optimally underreporting while on assistance is the same across regimes. We shall exploit this result repeatedly in what follows.

V. Revealed Preference Restrictions

The model above restricts how a woman responds to policy variation. That is, it rules out certain pairings of choices across the two policy regimes. These restrictions stem from simple revealed preference arguments. Specifically, if the utility of a woman's choice under AFDC is not lowered by the reform, she will either make the same choice under JF or select an alternative that the reform made more attractive.

A parsimonious approach to summarizing these restrictions leverages the fact that the JF reform improved (or worsened) the attractiveness of large collections of alternatives based on their implied earnings. To see this, recall that the JF reform altered the mapping between earnings and grant amounts and imposed more stringent work requirements on recipients with zero earnings. In what follows, we group labor supply alternatives into three broad categories based upon the earnings they generate. We then apply revealed preference arguments to rule out possible pairings of these broad categories of alternatives across policy regimes.[17] In Section X, we discuss what can (and cannot) be learned from working with finer earnings categories.

A. *Earnings Ranges*

Consider the following "coarsened" earnings variable $\tilde{E}_i$, defined by the relation

$$\tilde{E}_i \equiv \begin{cases} 0 & \text{if } E = 0 \\ 1 & \text{if } E \leq FPL_i \\ 2 & \text{if } E > FPL_i \end{cases}. \tag{7}$$

That is, $\tilde{E}_i$ indicates whether woman i works, and if so, whether her earnings make her ineligible for welfare assistance under JF.

The JF reform had qualitatively similar effects on the attractiveness of alternatives within each of these earnings ranges. To see this, note that the reform potentially reduced the attractiveness of not working ($\tilde{E}_i = 0$) while on welfare because of JF's more stringent work requirements. By contrast, the reform made earning positive amounts below the poverty line ($\tilde{E}_i = 1$) at least as attractive since a woman with earnings in this range is either off assistance, or on assistance and underreporting, or on assistance and truthfully reporting. In the first two circumstances, the utility value she attains is unaffected by the regime (recall the optimal reporting result

[17] The consideration of restrictions on pairings of broad ranges of alternatives is common in the theoretical revealed preference literature (see McFadden 2005 and Kitamura and Stoye 2013). This approach is to be distinguished from the common practice in structural labor supply papers of assuming labor supply choices are constrained to fall into a few data-driven categories such as "part-time" and "full-time" work (e.g., Hoynes 1996; Keane and Moffitt 1998; Blundell et al. 2013; Manski 2014). Here, we allow the choice set to vary across women in an unrestricted fashion by means of the heterogeneous offer set Θ_i.

described in Section IVB). In the third circumstance, the utility value she attains under JF is at least as high as that attainable under AFDC because of JF's enhanced earnings disregard. Finally, the reform had no effect on the utility of working at earnings levels above the poverty line ($\tilde{E}_i = 2$). This follows because a woman with earnings in range 2 is either off assistance or underreporting while on assistance. In both circumstances, the utility value she attains is unaffected by the regime.

Pairing the earning categories with the decision to participate in welfare and the underreporting decision yields seven earnings/participation/reporting combinations, which we henceforth refer to as *states*. The set of possible states is given by

$$\mathcal{S} \equiv \{0n, 1n, 2n, 0r, 1r, 1u, 2u\}.$$

The number associated with each state refers to the woman's earnings category while the letter describes her combined welfare participation and reporting decisions. Specifically, the letter n denotes welfare nonparticipation, r denotes welfare participation with truthful reporting of earnings ($E^r = E$), and u denotes welfare participation with underreporting of earnings ($E^r < E$). Note that state $0u$ is ruled out, as it is not meaningful to "underreport" zero earnings. Likewise, state $2r$ cannot occur under either JF or AFDC because of their respective eligibility rules (recall that we are considering a woman with $\overline{E}_i \leq FPL_i$).

B. *Allowed and Disallowed Responses*

Table 3 catalogs the possible pairings of states across the two policy regimes. Pairs of states labeled "No response" entail the same behavior under the two policy regimes. We term the remaining pairs either "disallowed" or "allowed" responses. The disallowed responses entail a change in behavior that is proscribed by the model. This occurs either because the change in behavior would entail an alternative that is dominated or because the change in behavior is incompatible with revealed preference. In Table 3, the disallowed responses are denoted with a "−." The allowed responses entail a change in behavior that is permitted by the model. These responses are represented by entries that describe the three margins along which behavior may change: welfare participation (welfare take-up or exit), labor supply (extensive versus intensive labor supply response), and reporting of earnings to the welfare agency (truthful reporting versus underreporting). We next describe the logic behind which responses are allowed and which are not. The online Appendix establishes formally that the restrictions in Table 3 are exhaustive.

Starting with the disallowed responses, a woman will not make a choice corresponding to state $1u$ under JF because underreporting is costly (A2) and earnings below the poverty line are fully disregarded. For this reason, the column of Table 3 pertaining to state $1u$ under JF is populated with "−" entries over a dark grayed background. The remaining prohibited responses stem from revealed preference arguments. By Assumptions (A1) and (A5), the JF reform may have made alternatives corresponding to state $1r$ more (but not less) attractive. Conversely, by Assumption (A4), the reform may have made alternatives corresponding to the state $0r$ less (but not more) attractive. Finally, the reform had no effect on the value of alternatives corresponding to the set $\mathcal{C}_0 \equiv \{0n, 1n, 2n, 1u, 2u\}$ by Assumption (A6) and optimal

Table 3—Allowed and Disallowed Responses

State under AFDC	State under Jobs First: $0n$	$1n$	$2n$	$0r$	$1r$	$1u$	$2u$
$0n$	No response	—	—	—	Extensive LS $(+)$ Take up welfare	—	—
$1n$	—	No response	—	—	Intensive LS $(+/0/-)$ Take up welfare	—	—
$2n$	—	—	No response	—	Intensive LS $(-)$ Take up welfare (Figure 5, panel B)	—	—
$0r$	No LS response Exit welfare	Extensive LS $(+)$ Exit welfare	Extensive LS $(+)$ Exit welfare	No response	Extensive LS $(+)$ (Figure 5, panel A)	—	Extensive LS $(+)$ Underreporting (Figure 6, panel B)
$1r$	—	—	—	—	Intensive LS $(+/0/-)$	—	—
$1u$	—	—	—	—	Intensive LS $(+/0/-)$ Truthful reporting	—	—
$2u$	—	—	—	—	Intensive LS $(-)$ Truthful reporting (Figure 6, panel A)	—	No response

Notes: This table catalogs the theoretically allowed response margins given the states that a woman may occupy under AFDC and Jobs First. A state is a pair of coarsened earnings (0 stands for zero earnings, 1 for positive earnings at or below the FPL, and 2 for earnings strictly above the FPL), and participation status in the relevant welfare assistance program along with an earnings reporting decision (n stands for "not on assistance," r for "on assistance and truthfully reporting earnings," and u for "on assistance and underreporting earnings"). The cells termed "No response" entail the same behavior under the two policy regimes. The cells containing "—" represent responses that are not allowed based on revealed-preference arguments derived from the model of Section IV. Specifically, (i) state $1u$ is unpopulated under JF ("—" in cells with a dark gray background) and (ii) a woman will not leave a state at least as attractive under JF as under AFDC for a state that is no more attractive under JF than under AFDC ("—" in cells with a light gray background). The remaining cells represent responses that are allowed by the model. Their content summarizes the three possible sorts of responses: (i) the labor supply (LS) response (intensive versus extensive and its sign: $+$ for increase, 0 for no change, and $-$ for decrease); (ii) the program participation response (take-up of versus exit from welfare assistance); and (iii) the reporting of earnings to the welfare agency margin (to truthfully report versus to underreport). See online Appendix for details.

reporting. Therefore, by revealed preference, a woman will not pair any of the states in $\mathcal{C}_{\succeq} \equiv \{1r\} \cup \mathcal{C}_0$ under AFDC with a (different) state in $\mathcal{C}_{\preceq} \equiv \{0r\} \cup \mathcal{C}_0$ under JF. This reasoning justifies the "$-$" entries in the cells with a light grayed background.

Proceeding now to responses that are allowed, consider first the extensive margin labor supply responses. A woman who, under AFDC, chooses not to work while off welfare (state $0n$) must face high welfare stigma, hassle, or underreporting costs since she is willing to forgo the full grant amount $\overline{G}_i$. Under JF, she may choose to work while on assistance and earn below the poverty line (state $1r$), as this option entails higher consumption than under AFDC. Next, a woman who, under AFDC, would participate in welfare without working (state $0r$), may respond to JF in several ways. Specifically, she may be induced to: (i) work while on welfare (state $1r$); (ii) leave welfare and earn less than the federal poverty line (state $1n$); (iii) leave welfare and earn more than the federal poverty line (state $2n$); (iv) remain on welfare and earn more than the federal poverty line (state $2u$); or (v) opt out of welfare (state $0n$). The first response can result from either the reduction in implicit tax rates on earnings or the increased hassle associated with JF. Sufficiently large fixed costs of work can enable the second, third, or fourth responses. A large increase in the hassle costs may induce the fifth response, in which case no labor supply response occurs.

Consider next the allowed intensive margin labor supply responses. The pairing of states $1n$, $1r$, or $1u$ under AFDC with state $1r$ under JF could entail intensive margin responses as a woman may adjust her earnings within region 1. A woman working on welfare under AFDC, and earning less than the poverty line, will face a reduction in her implicit tax rate under JF. Like any uncompensated increase in the wage, this change could lead to increases or decreases in the amount of work undertaken, but in either case will lead her to continue working on welfare. Likewise, a woman working off welfare under AFDC may choose to participate in JF which would offer an increase in income for the same amount of work. This may result in a reduction in earnings due to income effects. If the woman has high enough welfare stigma, she will not participate in welfare under either regime (i.e., she will pair state $1n$ with state $1n$). The pairing of either state $2n$ or $2u$ under AFDC with state $1r$ under JF also corresponds to an intensive margin response: the reform induces the woman to reduce her earnings below the poverty line.

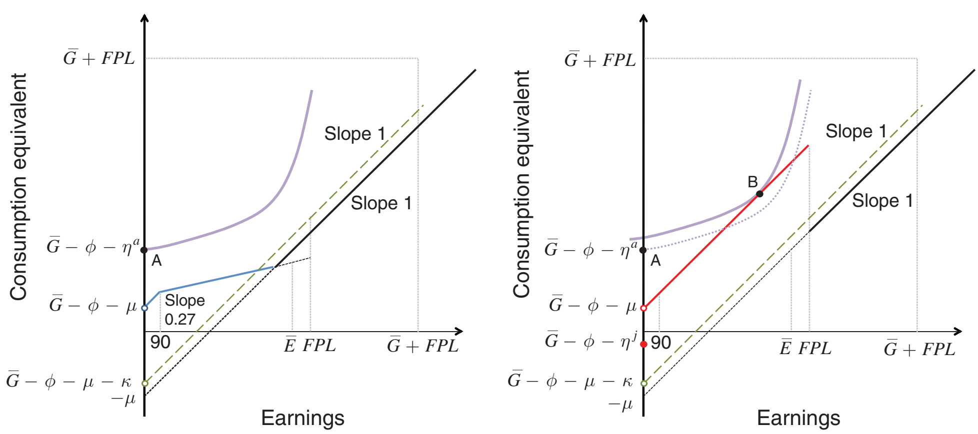
Some of the extensive and intensive margin labor supply responses above can be accompanied by an adjustment in reporting behavior. Specifically, the JF reform may induce a woman to truthfully report her earnings (pairing states $1u$ or $2u$ with state $1r$). Conversely, the reform may induce a woman to underreport her earnings (pairing state $0r$ with state $2u$). Thus, the JF reform may have mixed effects on reporting behavior.

C. *Graphical Examples*

Figures 5 and 6 illustrate some of the allowed responses listed above assuming a fixed wage rate and no labor market constraints (i.e., setting $K_i = \infty$). For convenience, both figures employ the monetized form of the utility function given in (5).

Figure 5 illustrates allowed responses that entail either an extensive margin or intensive margin labor supply adjustment. Specifically, panel A illustrates an extensive margin response, corresponding to pairing state $0r$ under AFDC with state $1r$ under JF. As depicted, the hassle costs η_i^a of not working under AFDC are much smaller than the corresponding costs η_i^j under JF. The fixed cost of work μ_i straddles the two hassle costs. In comparison with the fixed costs of work and hassle, the cost of underreporting κ_i is depicted as being quite large. The underreporting line is the same under AFDC and JF because under either regime a woman can secure the base grant by concealing her earnings. A woman with the configuration of preferences found in panel A would not work on welfare under AFDC (point A) but would take up work and truthfully report her earnings under JF (point B). Panel B illustrates the traditional opt-in response considered in the literature, corresponding to pairing state $2n$ under AFDC with state $1r$ under JF. As depicted, the hassle costs η_i^a of not working under AFDC are large but smaller than the corresponding costs η_i^j under JF. The fixed cost of work μ_i again straddles the two hassle costs. A woman with the configuration of costs and preferences found in panel B would earn above the

Panel A. From not working on assistance under AFDC to earning in range 1 under Jobs First



Panel B. From working in range 2 off assistance under AFDC to earning in range 1 under Jobs First

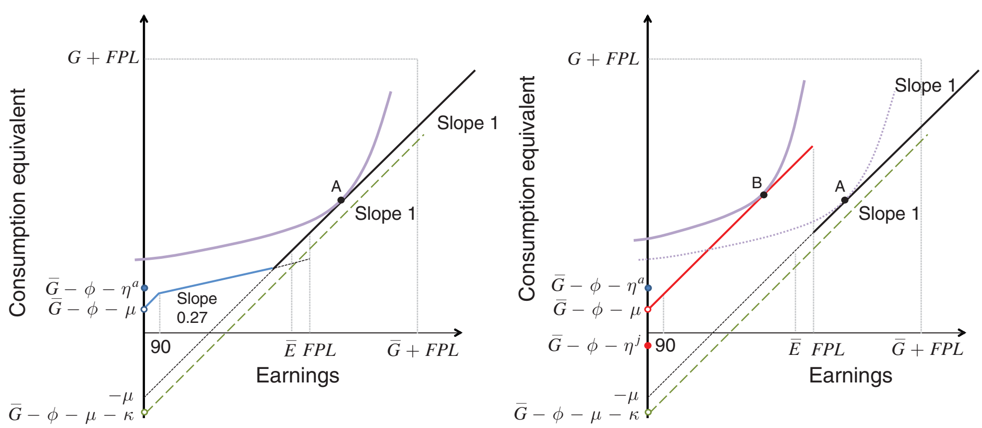


Figure 5. Extensive and Intensive Margin Responses to Reform

Notes: Panels A and B are drawn in the earnings (horizontal axis) and consumption equivalent (vertical axis) plane. The consumption equivalent equals earnings plus transfer income from welfare (if any) net of monetized hassle, stigma, work, and underreporting costs (if any: see text for details). At each level of earnings, the bold lines correspond to consumption either off welfare or on welfare with truthful reporting of earnings to the welfare agency. The dashed lines correspond to consumption on welfare with underreporting. Vertical lines represent the same earnings levels depicted in Figure 1: the fixed earning disregard under AFDC (\$90), the earnings level $\bar{E}$ at which welfare assistance is exhausted under AFDC, and the FPL. For clarity, the graphs assume away earnings constraints and use a fixed wage rate. Panel A depicts a scenario where the JF reform induces a woman who would participate in welfare and not work under AFDC (point A) to take up work and truthfully report her earnings under JF (point B)—an extensive margin response. Panel B depicts a scenario where the JF reform induces a woman to be off assistance and earn in range 2 (point A) to reduce her earnings to range 1 and take up assistance under JF (point B)—an intensive margin response.

poverty line off assistance under AFDC (point A) but would earn strictly below the poverty line on assistance under JF (point B).

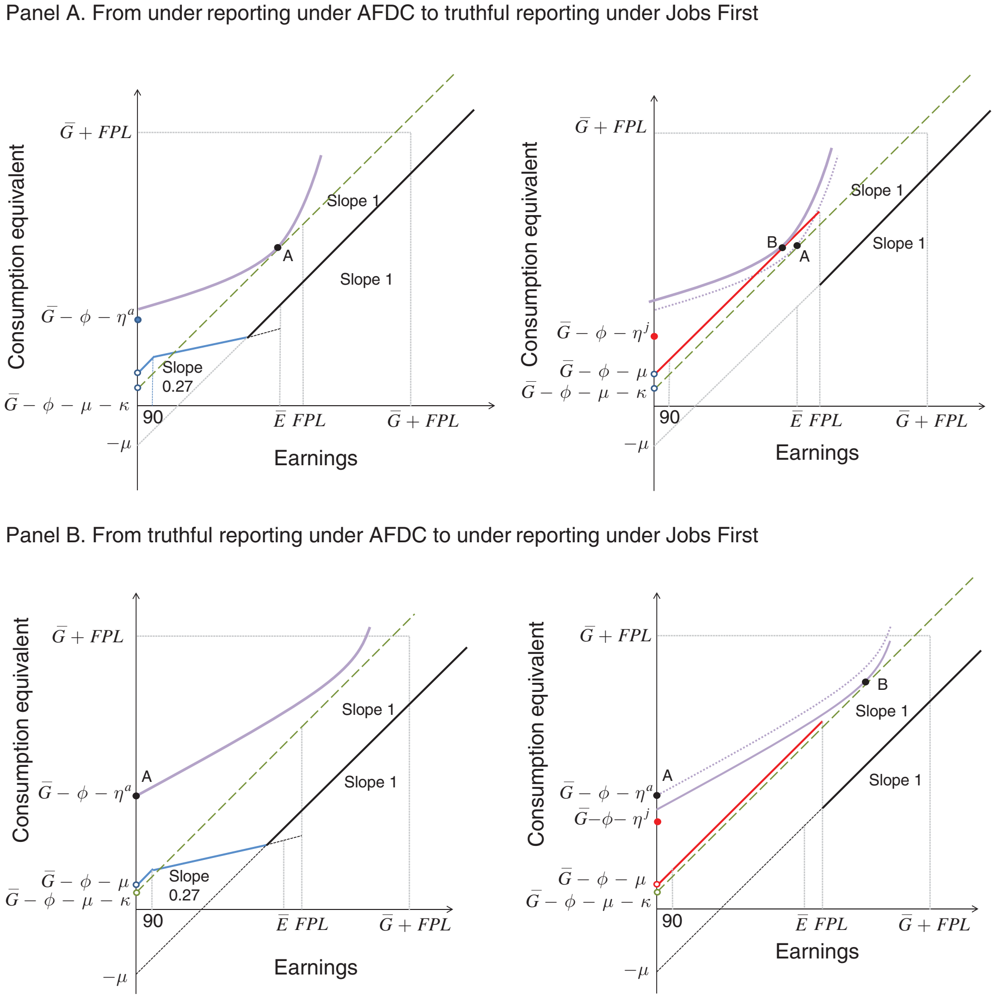


Figure 6. Earnings and Participation Choices with Underreporting

Notes: Panels A and B are drawn in the earnings (horizontal axis) and consumption equivalent (vertical axis) plane. The consumption equivalent equals earnings plus transfer income from welfare (if any) net of monetized hassle, stigma, work, and underreporting costs (if any: see text for details). At each level of earnings, the bold lines correspond to consumption either off welfare or on welfare with truthful reporting of earnings to the welfare agency. At each level of earnings, the dashed lines correspond to consumption on welfare with underreporting. Vertical lines represent the same earnings levels depicted in Figure 1: the fixed earning disregard under AFDC (\$90), the earnings level at $\bar{E}$ which welfare assistance is exhausted under AFDC, and the FPL. For clarity, the graphs assume away earnings constraints and use a fixed wage rate. Panel A depicts a scenario where the JF reform induces a woman who would participate in welfare, work, and underreport her earnings under AFDC (point A) to work and truthfully report her earnings under JF (point B) thanks to the 100 percent earning disregard under JF. Panel B depicts a scenario where the JF reform induces a woman who would participate in welfare without work under AFDC (point A) to work and underreport her earnings under JF (point B) to avoid the hassle cost under JF.

Figure 6 illustrates allowed responses that entail an adjustment in reporting behavior. As depicted, the hassle costs η_i^j of not working under JF are larger than the corresponding costs η_i^a under AFDC, but both are smaller than the fixed cost of work μ_i. In comparison with the fixed costs of work and hassle, the cost of underreporting κ_i

is relatively small. A woman with the configuration of preferences found in panel A of Figure 6 would work on welfare under AFDC but underreport her earnings (point A). However, under JF, she would truthfully report her earnings (point B), as the JF disregard reduces the return to underreporting. Hence, reform may induce a reduction in underreporting. By contrast, panel B of Figure 6 shows a scenario where the hassle effects of JF are larger, the costs of underreporting are smaller, and preferences over earnings are such that the disutility of work is lower. This woman would receive benefits without working (point A) under AFDC but, under JF, will choose to earn above the poverty line and underreport her earnings (point B) in order to maintain eligibility. This occurs because the JF work requirements remove point A from her budget set—such a woman has effectively been hassled off welfare into underreporting.

VI. Extensions

The model of Section IV pertains to a woman who, under AFDC, would have access to the reduced earnings disregards so that $\overline{E}_i \leq FPL_i$. Here we extend the model by considering a woman for whom $\overline{E}_i > FPL_i$. We show that, for such a woman, the set of theoretically allowable responses to the JF reform is expanded but that these additional responses are empirically irrelevant for the women in our analysis sample. Additionally, we summarize why the inclusion of food stamps, payroll taxes, and the EITC does not change our reasoning about the theoretically allowable effects of the JF reform with respect to welfare take-up and earnings. The important conclusion to be drawn from these two extensions is that the response margins cataloged in Section V constitute the full list of possible welfare participation and earning responses for the women in our sample.

A. *Unreduced Disregards*

A woman with break-even earnings level $\overline{E}_i > FPL_i$ may occupy state $2r$ because AFDC rules permit her to truthfully report earnings above the poverty line. This augments the set of possible responses to the JF reform since, as remarked above, state $2r$ is not permitted by the JF eligibility rules. In particular, access to unreduced disregards enables flows out of the labor force (i.e., pairing of state $2r$ under AFDC with state $0r$ or $0n$ under JF) provided that earning constraints are present (see the online Appendix for details).

While interesting, the analysis of these additional responses turns out to be purely pedagogical. The number of observations in our sample for which this sort of behavior could be present is bounded from above by the number of quarters in the AFDC sample where women earn more than the poverty line and receive a positive welfare transfer no larger than $G_i^a(FPL_i)$. In our data, there are only 3 case-quarters (out of 14,784) meeting these criteria, implying that such behavior is extremely rare.[18]

[18] This estimate is constructed as follows: for each AFDC sample woman and quarter, we determine the welfare transfer she would receive if her earnings equaled the (assistance unit size and quarter-specific) poverty line and if she had access to the unreduced fixed and proportional disregards. We round this amount to the nearest \$50 and denote it by $G_i^{a*}(FPL_i)$. Then, we count the number of quarterly observations in the AFDC sample associated with UI earnings above the poverty line and with quarterly welfare transfers no greater than $G_i^{a*}(FPL_i)$.

This should not be particularly surprising—if women have convex preferences they are unlikely to earn in the range $(FPL_i, \overline{E}_i]$ since AFDC benefit exhaustion induces a nonconvex kink in the budget set (Moffitt 1990). Moreover, even a mild welfare stigma could outweigh the relatively trivial amount of cash assistance available to women with earnings in this range. Whatever the explanation, the conclusion is the same: disregarding state $2r$ (and the related response margins) is empirically inconsequential.

B. *Food Stamps*, *the EITC*, *and Payroll Taxes*

In the online Appendix we develop an extended model where Food Stamps participation is introduced as an additional choice variable, so that a woman may be off assistance, on welfare only, on food stamps only, or on both welfare and Food Stamps.[19] Pairing the earnings categories in (7) with the decisions to participate in welfare and/or Food Stamps as well as the underreporting decision yields 16 states. Revealed preference arguments proscribe 190 out of the $16 \times 15 = 240$ atheoretically possible responses leaving us with 50 allowed responses. Crucially, none of the allowed responses involve pairing of earnings, welfare participation, and reporting alternatives prohibited by the model of Section IV. The reason for this convenient result is that, under JF, earnings up to the poverty line were disregarded in full for the determination of the Food Stamps grant *only conditional on joint take-up* of welfare. Thus, JF's impact on the Food Stamps program effectively amplifies the notch at the poverty line (recall Figure 2) and leaves the attractiveness of the *nonwelfare* assistance states unaffected. In summary, the restrictions in Table 3 hold with reference to both the welfare and Food Stamps components of the JF reform and given the tax system in place at the time of the reform.

VII. Identification of Response Probabilities

Table 3 summarizes the restrictions our model places on how a woman may respond to the JF reform. These restrictions are not directly testable because we cannot observe the same woman under both policy regimes at a given point in time. However, the experimental nature of our data allow us to compare groups of women with identical *distributions* of primitives who face different policy rules. In this section, we discuss how the individual level restrictions enumerated in Table 3 can be exploited to test the model and bound the frequency of adjustment to the JF reform along each allowable response margin.

A. *Population Heterogeneity*

We start by assuming the N women in our sample obey assumptions (A1)–(A6) and have primitives $\{\theta_i\}_{i=1}^{N}$ drawn independently from a distribution function $\Gamma_\theta(\cdot)$. We shall depart from much of the structural labor supply literature by leaving the

[19] This model allows for separate stigma effects for each combination of food stamps and welfare assistance. Underreporting costs also vary depending on the type of assistance. Filing for EITC is assumed invariant to the policy regime, and payroll and Medicare taxes are levied on earnings under both regimes.

distribution $\Gamma_\theta(\cdot)$ unrestricted save for the support limitations implied by assumptions (A1)–(A6), and the logical nonnegativity of hours and wage offers. Substantively, this formulation implies that preferences and constraints may vary freely across women, giving rise, for instance, to arbitrary correlations between tastes and offer sets. Such dependence poses difficult endogeneity problems bypassed in much of the recent literature on nonparametric identification of structural labor supply models, which typically treats wages (and policy rules) as exogenous (Manski 2014; Blomquist et al. 2014).

Because we allow for unrestricted heterogeneity across women, the right mix of preferences and offers can rationalize any distribution of choices under a given policy regime. However, as we show below, our theoretical restrictions do have empirical content when applied to the JF experiment.

B. *The Identification Problem*

Let S_i^a denote the "potential" state corresponding to woman i's choice under the AFDC regime and S_i^j the state corresponding to her choice under the JF regime. Our goal is to identify *response probabilities* of the form

$$\pi_{s^a, s^j} \equiv P\left(S_i^j = s^j \,|\, S_i^a = s^a\right),$$

for $\left(s^a, s^j\right) \in \mathcal{S} \times \mathcal{S}$, where $P(\cdot)$ is the probability measure induced by the distribution function $\Gamma_\theta(\cdot)$. These probabilities summarize the frequency of adjustment to the JF reform along specific labor supply and participation margins. For example, $\pi_{2n,1r}$ gives the proportion of those women who under AFDC would earn above the federal poverty line while off assistance that would work on welfare under JF—that is, the share of high earning women who opt in to welfare in response to reform.

Let T_i denote the treatment regime to which woman i is assigned and $S_i \equiv \mathbf{1}[T_i = j]\, S_i^j + \mathbf{1}[T_i = a]\, S_i^a$ her realized state. Random assignment ensures that her potential states are independent of the policy regime to which she is assigned. Formally,

$$T_i \perp \left(S_i^a, S_i^j\right), \tag{8}$$

where the symbol $\perp$ denotes independence. The condition above implies that, for every $s \in \mathcal{S}$ and $t \in \{a, j\}$, $P(S_i = s \,|\, T_i = t) = P(S_i^t = s) \equiv q_s^t$, which is the well-known result that experimental variation identifies the marginal distributions of potential outcomes.

Unfortunately, experimental variation is not sufficient to identify the response probabilities $\{\pi_{s^a, s^j}\}$. To see this, observe that by the law of total probability, the marginal distributions of potential states are linked by the relation

$$\mathbf{q}^j = \mathbf{\Pi}'\mathbf{q}^a, \tag{9}$$

where $\mathbf{q}^t \equiv \left[q_{0n}^t, q_{1n}^t, q_{2n}^t, q_{0r}^t, q_{1r}^t, q_{1u}^t, q_{2u}^t\right]'$ for $t \in \{a, j\}$ and the 7×7 matrix $\mathbf{\Pi}$ is composed of unknown response probabilities. Supposing for the moment that we know the vectors $(\mathbf{q}^a, \mathbf{q}^j)$ with certainty, the system in (9) consists of 7 equations (1 of which is redundant) and $7 \times 6 = 42$ unknown independent response probabilities. Clearly, the response probabilities are heavily underidentified. As we show next, our model dramatically reduces the degree of underidentification present.

C. *Unrestricted Response Probabilities*

The revealed preference arguments developed in Section IV imply that only 10 out of the 42 possible response margins cataloged in Table 3 are allowed. Accordingly, only 10 of the 42 response probabilities in the matrix $\mathbf{\Pi}$ are not restricted to equal zero. Furthermore, $\pi_{1u,1r}$ equals 1 because no woman pairs state $1u$ under AFDC with any state but $1r$ under JF. Hence, there are 9 free response probabilities, which we collect into the vector,

$$(10) \qquad \boldsymbol{\pi} \equiv \left[\pi_{0n,1r}, \pi_{0r,0n}, \pi_{2n,1r}, \pi_{0r,2n}, \pi_{0r,1r}, \pi_{0r,1n}, \pi_{1n,1r}, \pi_{0r,2u}, \pi_{2u,1r}\right]'.$$

Note that rank reversals in earnings may result from the allowed responses to reform. For example, some women who do not work under AFDC may earn in range 2 under JF $(\pi_{0r,2n} > 0$ or $\pi_{0r,2u} > 0)$ and thereby "leap over" their peers who earn in range 1 of the earnings distribution under either regime.[20]

Recall that we have only seven equations to discipline the nine free response probabilities, which necessitates a partial identification analysis of $\boldsymbol{\pi}$. Moreover, because we do not directly observe under-reporting behavior, we cannot distinguish between states $1u$ and $1r$, making the vectors $(\mathbf{q}^a, \mathbf{q}^j)$ themselves underidentified. We address both of these concerns below.

D. *Observable States*

Our data do not allow us to measure reporting decisions other than by contrasting a woman's administrative earnings with the eligible maximum. Hence, states $1u$ and $1r$ are not empirically distinguishable. Accordingly, we define a function $g : \mathcal{S} \rightarrow \tilde{\mathcal{S}}$ that reduces the *latent* states $\mathcal{S}$ to *observable* states $\tilde{\mathcal{S}}$ that can be measured in our data. Formally,

$$g(s) \equiv \begin{cases} s & \text{if } s \in \{0n, 1n, 2n\} \\ 0p & \text{if } s = 0r \\ 1p & \text{if } s \in \{1u, 1r\} \\ 2p & \text{if } s = 2u \end{cases}.$$

As before, the number of each state refers to the woman's earnings category and the letter n refers to welfare nonparticipation. The letter p denotes welfare

[20] Many additional sorts of rank reversals are possible in this framework. For example, additional reversals occur in the case above where $\pi_{0r,2n} > 0$ or $\pi_{0r,2u} > 0$ if some women who would earn in range 2 under AFDC earn in range 1 under JF $(\pi_{2n,1r} > 0$ or $\pi_{2u,1r} > 0)$. One also expects reversals to be common among women who do not work under AFDC but earn positive amounts under JF, since the amounts earned under JF are likely to be above the lowest earnings levels of women who would have worked even in the absence of reform.

participation, which is directly observable. Note that state $2p$ can only be occupied via underreporting.

Let $\tilde{S}_i^t$ denote the potential observable state of a woman whose latent potential state under policy regime t is S_i^t, that is, $\tilde{S}_i^t \equiv g(S_i^t)$ for $t \in \{a,j\}$. Also, define the probability of occupying state $\tilde{s} \in \tilde{\mathcal{S}}$ under policy regime t as $p_{\tilde{s}}^t \equiv P\left(\tilde{S}_i^t = \tilde{s}\right) = \sum_{s:\,\tilde{s}=g(s)} q_s^t$. Finally, denote the vectors of *observable* state probabilities as $\mathbf{p}^t \equiv \left[p_{0n}^t, p_{1n}^t, p_{2n}^t, p_{0p}^t, p_{1p}^t, p_{2p}^t\right]'$ for $t \in \{a,j\}$. We are now ready to discuss identification of the nine free response probabilities appearing in (10) based on the regime-specific state distributions $\mathbf{p}^a$ and $\mathbf{p}^j$.

E. *Identified Set*

Integrating the unobserved states out of (9) yields a system of six equations, one of which is redundant given that state probabilities sum to one in each policy regime. The five nonredundant equations can be given an intuitive representation as

$$
\begin{aligned}
(11)\qquad p_{0n}^j - p_{0n}^a &= -p_{0n}^a \pi_{0n,1r} + p_{0p}^a \pi_{0r,0n} \\
p_{1n}^j - p_{1n}^a &= -p_{1n}^a \pi_{1n,1r} + p_{0p}^a \pi_{0r,1n} \\
p_{2n}^j - p_{2n}^a &= -p_{2n}^a \pi_{2n,1r} + p_{0p}^a \pi_{0r,2n} \\
p_{0p}^j - p_{0p}^a &= -p_{0p}^a \left(\pi_{0r,1n} + \pi_{0r,1r} + \pi_{0r,2u} + \pi_{0r,2n} + \pi_{0r,0n}\right) \\
p_{2p}^j - p_{2p}^a &= p_{0p}^a \pi_{0r,2u} - p_{2p}^a \pi_{2u,1r}.
\end{aligned}
$$

The left-hand side of (11) catalogs the experimental impacts of the JF reform on the observable state probabilities. The right-hand side rationalizes these impacts in terms of "flows" into and out of each state as allowed by the model. The identifying power of the model derives from the fact that only a handful of response probabilities appear in each equation. The *identified set* Ξ of response probabilities consists of the set of vectors $\boldsymbol{\pi}$ obeying (11) that satisfy the usual adding up and nonnegativity conditions of probability distributions.[21]

F. *Testable Restrictions*

As we show in the online Appendix, System (11) implies 16 inequality restrictions. These restrictions exhaust the predictions of our model for the distribution of observed states $\left(\mathbf{p}^a, \mathbf{p}^j\right)$. As argued above, the restrictions pertain exclusively to the impact $(\mathbf{p}^j - \mathbf{p}^a)$ of the JF reform on state probabilities, as opposed to the cross-sectional distributions of states within a regime. Violation of any of these

[21] Here this means that $\pi \in [0,1]^9$ and $\pi_{0r,1n} + \pi_{0r,1r} + \pi_{0r,2u} + \pi_{0r,2n} + \pi_{0r,0n} \leq 1$.

inequalities would imply that our framework fails to allow for a response actually present in the data. To conserve space, we list the 16 inequality restrictions in the online Appendix. Here we report five of them that are particularly intuitive:

$$\left(p_{0p}^{j} - p_{0p}^{a}\right) \leq 0 \tag{12a}$$

$$\left(p_{0p}^{j} - p_{0p}^{a}\right) + \left(p_{0n}^{j} - p_{0n}^{a}\right) \leq 0 \tag{12b}$$

$$\left(p_{0p}^{j} - p_{0p}^{a}\right) + \left(p_{2n}^{j} - p_{2n}^{a}\right) + \left(p_{0n}^{j} - p_{0n}^{a}\right) + \left(p_{1n}^{j} - p_{1n}^{a}\right) \leq 0 \tag{12c}$$

$$\left(p_{0p}^{j} - p_{0p}^{a}\right) + \left(p_{2n}^{j} - p_{2n}^{a}\right) + \left(p_{0n}^{j} - p_{0n}^{a}\right) + \left(p_{2p}^{j} - p_{2p}^{a}\right) \leq 0 \tag{12d}$$

$$\left(p_{0p}^{j}-p_{0p}^{a}\right) + \left(p_{2n}^{j}-p_{2n}^{a}\right) + \left(p_{0n}^{j}-p_{0n}^{a}\right) + \left(p_{2p}^{j}-p_{2p}^{a}\right) + \left(p_{1n}^{j}-p_{1n}^{a}\right) \leq 0. \tag{12e}$$

These restrictions state that the JF reform must (weakly): lower the fraction of women on assistance and not working (12a); raise the fraction of women working (12b); raise the fraction of women who work and receive assistance (12c); raise the fraction of women with earnings in range 1 (12d); and raise the fraction of women who receive assistance and have earnings in range 1 (12e).

G. *Further Structuring Preferences*

As an illustration of the identifying power of further structuring preferences, we also consider the monetized form of the utility function given in (5). In the online Appendix we show that with this restricted specification, the choice of $0r$ under AFDC by woman i reveals that her stigma cost ϕ_i is below the base grant amount $\overline{G}_i$. This, in turn, implies that state $1n$ is dominated by state $1r$ under JF. Hence, no woman pairs state $0r$ under AFDC with state $1n$ under JF. Accordingly, $\pi_{0r,\,1n} = 0$ which reduces the number of unknown response probabilities to 8. Imposing this restriction on system (11) reveals that the second equation uniquely identifies the response probability $\pi_{1n,\,1r}$. Intuitively, when $\pi_{0r,\,1n} = 0$, there is a "flow" into but no "flow" out of state $1n$. Furthermore, this version of the model implies that the JF reform must (weakly) reduce the fraction of working women off assistance and earning in range 1, formally:

$$p_{1n}^{a} - p_{1n}^{j} \geq 0. \tag{13}$$

VIII. Bounds on Response Probabilities

Subject to the restrictions above holding, we can derive bounds on the nine response probabilities that are set-identified in our baseline model. The upper and lower bounds on each of the response probabilities correspond to vertices of the identified set Ξ. These vertices can be represented as the solution to a pair of linear programming problems of the form

$$\max_{\boldsymbol{\pi}} \boldsymbol{\pi}'\boldsymbol{\lambda} \text{ subject to } \boldsymbol{\pi} \in \Xi, \tag{14}$$

where the layout of $\boldsymbol{\pi}$ was given in (10). For example, solving the problem above for $\boldsymbol{\lambda} = [0,0,1,0,0,0,0,0,0]'$ yields the upper bound on $\pi_{2n,1r}$, while choosing $\boldsymbol{\lambda} = [0,0,-1,0,0,0,0,0,0]'$ yields the lower bound.

A. *Composite Margins*

We can also use this representation to derive bounds on linear combinations of the response probabilities. We consider the probabilities of adjusting along four "composite" margins:

$$\pi_{0r,n} \equiv \pi_{0r,0n} + \pi_{0r,2n} + \pi_{0r,1n},$$

$$\pi_{p,n} \equiv \frac{p^a_{0p}}{p^a_{0p} + p^a_{1p} + p^a_{2p}}\left(\pi_{0r,0n} + \pi_{0r,2n} + \pi_{0r,1n}\right),$$

$$\pi_{n,p} \equiv \frac{p^a_{0n}\pi_{0n,1r} + \pi_{1n,1r}p^a_{1n} + \pi_{2n,1r}p^a_{2n}}{p^a_{0n} + p^a_{1n} + p^a_{2n}},$$

$$\pi_{0,1+} \equiv \frac{p^a_{0p}\left(\pi_{0r,1r} + \pi_{0r,2n} + \pi_{0r,2u} + \pi_{0r,1n}\right) + p^a_{0n}\pi_{0n,1r}}{p^a_{0p} + p^a_{0n}}.$$

The first composite response probability gives the fraction of women who would claim benefits without working under AFDC that are induced to get off welfare under JF (denoted $\pi_{0r,n}$). Upper and lower bounds for this response probability can be had by solving (14) with $\boldsymbol{\lambda} = [0,1,0,1,0,1,0,0,0]$ and $[0,-1,0,-1,0,-1,0,0,0]$ respectively. We also examine the fraction of all women who would participate in welfare under AFDC that are induced to leave welfare under JF (denoted $\pi_{p,n}$), the fraction of women who are induced to take up welfare under JF (denoted $\pi_{n,p}$), and the fraction of women who are induced by JF to work (denoted $\pi_{0,1+}$). Because no woman who would work under AFDC will choose not to work under JF, this last fraction is point identified by the proportional reduction in the fraction of women not working under JF relative to AFDC.

B. *Analytic Expressions*

It is useful for conducting inference to obtain analytic expressions for the bounds as a function of the regime-specific marginal distributions $(\mathbf{p}^a, \mathbf{p}^j)$. We accomplished this by solving the relevant linear programming problems by hand. The resulting expressions are listed in the online Appendix. An example is given by the bounds on the opt-in probability $\pi_{2n,1r}$ which take the form

$$(15)\quad \max\left\{0, \frac{p^a_{2n} - p^j_{2n}}{p^a_{2n}}\right\} \leq \pi_{2n,1r}$$

$$\leq \min\left\{\begin{array}{c} 1, \\ \dfrac{\left(p^a_{2n} - p^j_{2n}\right) + \left(p^a_{0p} - p^j_{0p}\right)}{p^a_{2n}}, \\ \dfrac{\left(p^a_{2n} - p^j_{2n}\right) + \left(p^a_{0p} - p^j_{0p}\right) + \left(p^a_{0n} - p^j_{0n}\right)}{p^a_{2n}}, \\ \dfrac{\left(p^a_{2n} - p^j_{2n}\right) + \left(p^a_{0p} - p^j_{0p}\right) + \left(p^a_{2p} - p^j_{2p}\right)}{p^a_{2n}}, \\ \dfrac{\left(p^a_{2n} - p^j_{2n}\right) + \left(p^a_{0p} - p^j_{0p}\right) + \left(p^a_{1n} - p^j_{1n}\right)}{p^a_{2n}}, \\ \dfrac{\left(p^a_{2n} - p^j_{2n}\right) + \left(p^a_{0p} - p^j_{0p}\right) + \left(p^a_{0n} - p^j_{0n}\right) + \left(p^a_{2p} - p^j_{2p}\right)}{p^a_{2n}}, \\ \dfrac{\left(p^a_{2n} - p^j_{2n}\right) + \left(p^a_{0p} - p^j_{0p}\right) + \left(p^a_{0n} - p^j_{0n}\right) + \left(p^a_{1n} - p^j_{1n}\right)}{p^a_{2n}}, \\ \dfrac{\left(p^a_{2n} - p^j_{2n}\right) + \left(p^a_{0p} - p^j_{0p}\right) + \left(p^a_{2p} - p^j_{2p}\right) + \left(p^a_{1n} - p^j_{1n}\right)}{p^a_{2n}}, \\ \dfrac{\left(p^a_{2n} - p^j_{2n}\right) + \left(p^a_{0p} - p^j_{0p}\right) + \left(p^a_{0n} - p^j_{0n}\right) + \left(p^a_{2p} - p^j_{2p}\right) + \left(p^a_{1n} - p^j_{1n}\right)}{p^a_{2n}} \end{array}\right\}.$$

Note that there are two possible solutions for the lower bound, one of which is zero.[22] This turns out to be a generic feature of the lower bounds for each of the nine set-identified response probabilities. Here, the lower bound reflects the model's restriction that reform-induced reductions in the fraction of women occupying state $2n$ can only be rationalized via opt-in behavior (when positive, $\frac{p^a_{2n} - p^j_{2n}}{p^a_{2n}}$ gives the proportional reduction in the fraction of women choosing state $2n$ under JF relative to AFDC). The upper bound on $\pi_{2n,1r}$ admits nine possible solutions, corresponding to settings where opt-in responses are accompanied by other responses that move probability mass to state $1r$. The upper bounds on the remaining response probabilities can have fewer or more solutions.

[22] From (11), the fraction of people occupying state $2n$ under JF may differ from that under AFDC because of an "in-flow" from state $0r$ (represented by $p^a_{0p}\pi_{0r,2n}$) or because of an "out-flow" to state $1r$ (represented by $-p^a_{2n}\pi_{2n,1r}$). If $p^a_{2n} - p^j_{2n} \leq 0$, the in-flow from state $0r$ must be at least as large as the out-flow to state $1r$. But this latter quantity may be zero, in which case the lower bound on $\pi_{2n,1r}$ is zero. If $p^a_{2n} - p^j_{2n} > 0$, the in-flow from state $0r$ can at most equal the out-flow to state $1r$, in which case this latter quantity must be at least $p^a_{2n} - p^j_{2n}$. Accordingly, the lower bound on $\pi_{2n,1r}$ is the π that solves $p^a_{2n} - p^j_{2n} = p^a_{2n}\pi$, namely $\pi = \frac{p^a_{2n} - p^j_{2n}}{p^a_{2n}}$.

C. *Estimation and Inference*

Consistent estimators of the upper and lower bounds of interest can be had by using sample analogues of the marginal probabilities and computing the relevant $\min\{\cdot\}$ and $\max\{\cdot\}$ expressions. For example, working off of equation (15), we use $\max\left\{0, \frac{\hat{p}_{2n}^{a} - \hat{p}_{2n}^{j}}{\hat{p}_{2n}^{a}}\right\}$ as an estimator of the lower bound for $\pi_{2n,1r}$.

Inference is complicated by the fact that the limit distribution of the upper and lower bounds depends upon uncertainty in which of the constraints in (14) bind—i.e., in which of the bound solutions is relevant. As discussed by Andrews and Han (2009), bootstrapping the empirical $\min\{\cdot\}$ and $\max\{\cdot\}$ of the sample analogues of the bound solutions will fail to capture the sampling uncertainty in the bounds, particularly when many constraints are close to binding. Several approaches to this problem have been proposed that involve conducting pretests for which population constraints are binding (e.g., Andrews and Barwick 2012). We take an alternative approach to inference that is simple to implement, avoids the choice of tuning parameters, and is consistent regardless of the constraints that bind. Specifically, we report two sets of confidence intervals, one "naïve" interval that will under-cover when many constraints are close to binding and one "conservative" interval that will cover regardless of which constraints bind (details on both are provided in the online Appendix).

The procedure for constructing the "naïve" confidence interval ignores the uncertainty in which constraints bind—that is, it assumes the bound solution that appears relevant given the sample analogues binds with probability 1. In such a case, results from Imbens and Manski (2002) imply a pointwise 95 percent confidence interval for the parameter in question can be constructed by extending the upper and lower bounds by $1.65\hat{\sigma}$, where $\hat{\sigma}$ is a nonparametric bootstrap estimate of the standard error of the sample moment used to define the relevant bound. These confidence intervals will provide valid inferences only if no other constraints are close to binding.

Our second procedure, which is also based on the bootstrap, assumes that all bound solutions are identical, in which case sampling uncertainty in all of the solution estimates affects the composite bound. As we demonstrate in the online Appendix, this procedure, which can be thought of as an asymptotically conservative version of the method of Chernozhukov, Lee, and Rosen (2013), covers the parameter under study with asymptotic probability greater than or equal to 95 percent regardless of which solutions bind. The lower limit of the resulting "conservative" confidence interval coincides with that of the naïve confidence interval because sampling uncertainty only affects one of the bound solutions in the $\max\{\cdot\}$ operator. However, the upper limit of the conservative confidence interval generally exceeds that from the naïve confidence interval, often by a substantial amount. Hence, the conservative confidence interval provides fully robust coverage, at the possible expense of reductions in asymptotic power.

D. *Panel Data*

Thus far, we have abstracted from the panel nature of our data and presumed access to only a single cross section of choices. In practice, we follow BGH in conducting a pooled analysis using all person-quarter observations in the seven quarters following random assignment. We treat all person-quarter observations as potentially separate decision makers in order to accommodate the likely possibility that the primitives θ_i governing a woman's behavior vary over time. For example, among women in the AFDC control group who chose state $0r$ in the quarter prior to random assignment, more than 20 percent had switched states by the first quarter after random assignment, which implies that either their preferences or constraints must have changed.[23] Without additional assumptions restricting this evolution, our (static) model has no special implications for sequences of choices over time as opposed to choices made by different women. However, the bootstrapping approach employed to obtain confidence intervals for the bounds accounts for the fact that the observations contributed by a woman over time may be dependent. Specifically, it resamples a woman's entire profile of choices made during the first seven quarters after random assignment.

IX. Results

Table 4 reports the estimated probabilities of occupying the six observable earnings and welfare participation states under each policy regime in the seven quarters after random assignment.[24] The 16 testable restrictions of our baseline model, as well as the additional restriction (13) associated with the monetized form of utility, are satisfied by the point estimates. There is a small but statistically significant increase in the fraction of quarters on welfare with earnings above the quarterly poverty line indicating that, on net, JF induced more women to underreport earnings than it induced to truthfully report them.

Table 5 provides estimates of the response probabilities that rationalize the impacts in Table 4. Panel A of the table reports estimates obtained under the general specification of preferences given in (4), while panel B reports estimates obtained under the monetized specification of preferences given in (5).

Starting with panel A, our most important finding is that the JF reform induced a substantial opt-in response among women who would have otherwise worked off welfare at earning levels above the poverty line. The estimated bounds imply that $\pi_{2n,1r} \geq 0.28$. That is, at least 28 percent of those women with ineligible earnings under AFDC decided to work at eligible levels under JF and participate in welfare—an intensive margin labor supply response. Accounting for sampling uncertainty in the bounds extends this lower limit to 20 percent, which is still quite substantial. The upper bounds for this parameter are not informative leading us to conclude that the interval $[0.20, 1]$ covers the true opt-in probability with 95 percent probability. We also find suggestive evidence of a second opt-in effect from

[23] The details of this exercise are in online Appendix Table A5 which provides the distribution of states occupied in quarters 1 through 7 among the subsample of woman assigned to AFDC who chose state $0r$ in the quarter prior to random assignment.

[24] We discard from our sample all quarters in which a woman's welfare participation status varies from month to month as it would be impossible to infer reliably whether such a women earned above the poverty line in the months when she was on welfare. This selection could confound the experimental impacts reported in Table 4 if the experiment influenced the probability of selection. However, we find that after adjusting for baseline covariates via a linear probability model, the frequency of these "mixed" quarters is roughly the same in the AFDC and JF groups: the estimated impact of JF on the probability of a quarter being mixed is 0.0063 (SE $=$ 0.0034). Hence, we interpret the impacts reported in Table 4 as average treatment effects on "unmixed" quarters.

Table 4—Probability of Earnings/Participation States

	Overall			Overall – adjusted		
	Jobs First	AFDC	Difference	Jobs First	AFDC	Difference
Pr(State=0n)	0.127	0.136	−0.009	0.128	0.135	−0.007
				(0.005)	(0.006)	(0.008)
Pr(State=1n)	0.076	0.130	−0.055	0.078	0.126	−0.048
				(0.004)	(0.005)	(0.006)
Pr(State=2n)	0.068	0.099	−0.031	0.069	0.096	−0.027
				(0.004)	(0.005)	(0.006)
Pr(State=0p)	0.366	0.440	−0.074	0.359	0.449	−0.090
				(0.008)	(0.008)	(0.010)
Pr(State=1p)	0.342	0.185	0.157	0.343	0.184	0.159
				(0.007)	(0.005)	(0.009)
Pr(State=2p)	0.022	0.009	0.013	0.023	0.009	0.014
				(0.002)	(0.001)	(0.002)
Quarterly observations	16,226	16,268		16,226	16,268	

Notes: Sample covers quarters 1–7 post-random assignment during which individual is either always on or always off welfare. Sample cases with kidcount missing are excluded. Number of state refers to earnings level, with 0 indicating no earnings, 1 indicating earnings below 3 times the monthly FPL, and 2 indicating earnings above 3FPL. n indicates welfare nonparticipation throughout the quarter while p indicates welfare participation throughout the quarter. Poverty line computed under assumption Assistance Unit size is one greater than amount implied by baseline kidcount variable. Adjusted probabilities are computed via propensity score reweighting. Standard errors computed using 1,000 block bootstrap replications (resampling at case level).

nonparticipation, this time entailing an extensive margin labor supply response. Although the sample bounds imply $\pi_{0n,1r} \in [0.06, 0.62]$, uncertainty in the bounds prevents us from rejecting the null that this response probability is actually zero.

We find a small but significant underreporting response attributable to the hassle effects of JF. A conservative 95 percent confidence interval for $\pi_{0r,2u}$ is $[0.02, 0.10]$. Thus, JF induced at least one subpopulation to underreport earnings. The finding that both $\pi_{0r,2u}$ and $\pi_{2n,1r}$ are significantly positive indicates that earnings rank reversals occurred in response to reform. JF also had a strong effect on entry into the program by the working poor. The bootstrap confidence interval for $\pi_{1n,1r}$ indicates that at least 32 percent of the women who would have worked off welfare under AFDC at earnings levels below the poverty line were induced to participate in JF at eligible earning levels.

The remaining response probabilities $(\pi_{0r,0n}, \pi_{0r,2n}, \pi_{0r,1n}, \pi_{0r,1r}, \pi_{2u,1r})$ each have zero lower bounds. However, we can reject the null that they are jointly zero. From (11) such a joint restriction implies $p_{0p}^{j} - p_{0p}^{a} = -\left(p_{2p}^{j} - p_{2p}^{a}\right)$, which is easily rejected in our data. Thus, at least some of these margins of adjustment are present. Among the probabilities in question, the candidate that seems most likely to be positive is $\pi_{0r,1r}$ which is the extensive margin response through which welfare reform has traditionally been assumed to operate.

The last four rows in panel A of Table 5 report the estimated bounds and corresponding confidence intervals for the composite margins described in Section VII. First is the probability $\pi_{0,1+}$ that a woman responds along the extensive margin from nonwork to work. A 95 percent CI for this probability is $[0.14, 0.20]$. Thus, JF

Table 5—Point and Set-Identified Response Probabilities

Response type	State occupied under AFDC	JF	Symbol	Estimate	SE	95 percent CI (naïve)	(conservative)
Panel A. General specification of preferences							
Detailed	$0n$	$1r$	$\pi_{0n,1r}$	{0.055, 0.620}		[0.000, 0.739]	[0.000, 0.775]
	$1n$	$1r$	$\pi_{1n,1r}$	{0.382, 0.987}		[0.317, 1.000]	[0.317, 1.000]
	$2n$	$1r$	$\pi_{2n,1r}$	{0.280, 1.000}		[0.198, 1.000]	[0.198, 1.000]
	$0r$	$0n$	$\pi_{0r,0n}$	{0.000, 0.170}		[0.000, 0.205]	[0.000, 0.216]
	″	$1n$	$\pi_{0r,1n}$	{0.000, 0.170}		[0.000, 0.211]	[0.000, 0.215]
	″	$2n$	$\pi_{0r,2n}$	{0.000, 0.154}		[0.000, 0.171]	[0.000, 0.225]
	″	$1r$	$\pi_{0r,1r}$	{0.000, 0.170}		[0.000, 0.205]	[0.000, 0.219]
	″	$2u$	$\pi_{0r,2u}$	{0.031, 0.051}		[0.022, 0.058]	[0.022, 0.099]
	$2u$	$1r$	$\pi_{2u,1r}$	{0.000, 1.000}		[0.000, 1.000]	[0.000, 1.000]
Composite	Not working	Working	$\pi_{0,1+}$	0.167	0.016	[0.136, 0.198]	[0.136, 0.198]
	Off welfare	On welfare	$\pi_{n,p}$	{0.231, 0.445}		[0.189, 0.484]	[0.189, 0.486]
	On welfare	Off welfare	$\pi_{p,n}$	{0.000, 0.119}		[0.000, 0.144]	[0.000, 0.148]
	On welfare, not working	Off welfare	$\pi_{0r,n}$	{0.000, 0.170}		[0.000, 0.205]	[0.000, 0.212]
Panel B. Restricted specification of preferences							
Detailed	$0r$	$1n$	$\pi_{0r,1n}$	0			
	$1n$	$1r$	$\pi_{1n,1r}$	0.382	0.039	[0.305, 0.459]	[0.305, 0.459]

Notes: Number of state refers to earnings level, with 0 indicating no earnings, 1 indicating earnings below three times the monthly FPL, 2 indicating earnings above three times the monthly FPL, and 1+ indicating positive earnings. n indicates welfare nonparticipation, r indicates welfare participation with truthful reporting of earnings, u indicates welfare participation with underreporting of earnings, p indicates welfare participation (irrespective of reporting). Composite response probabilities are linear combination of the detailed response probabilities (see Section VIII for the exact expressions). Estimates inferred from probabilities in Table 4: see text for formulas. Numbers in braces are estimated upper and lower bounds, numbers in brackets are 95 percent confidence intervals. Naïve 95 percent confidence interval (CI) ignores uncertainty in which moment inequalities bind. Conservative 95 percent confidence interval assumes all constraints bind. See online Appendix for details. Panel A refers to the general specification of preferences (expression (4) in the paper). Panel B refers to the restricted specification (expression (5) in the paper). Panel B omits all response probabilities whose estimates are the same as in panel A.

induced a substantial fraction of women who would not have worked under AFDC to obtain employment under JF.

The confidence interval on the fraction $\pi_{n,p}$ of women induced to take up welfare by JF is relatively tight. Although JF unambiguously increased the fraction of women on welfare, our model suggests some women may also have been induced to leave welfare, breaking point identification of this margin. According to our conservative inference procedure, at least 19 percent (and at most 49 percent) of women off welfare under AFDC were induced to claim benefits under JF. Conversely, the fraction $\pi_{p,n}$ of women induced by JF to leave welfare is estimated to be at most 15 percent.

Finally, we cannot reject the null hypothesis that JF failed to induce any of the women who would have not worked while claiming AFDC benefits to leave welfare under JF, as the lower bound for the response probability $\pi_{0r,n}$ is zero. We are however able to conclude that at most 21 percent of such women left welfare, which may limit concerns that the JF reforms pushed a large fraction of women potentially unable to work off assistance.

Consider now panel B, which reports results when utility is assumed to be of the monetized form given in (5). Here, the response probability $\pi_{0r,1n}$ is constrained to equal zero which renders $\pi_{1n,1r}$ point identified. According to these estimates, the JF reform had a strong effect on entry into the program by the working poor. The bootstrap confidence interval for $\pi_{1n,1r}$ indicates that between 31 percent and 46 percent of the women who would have worked off welfare under AFDC at earnings levels below the poverty line were induced to participate in JF at eligible earning levels. The estimates of the remaining response probabilities and composite margins are omitted because they are the same as in panel A.[25]

X. Robustness

We now discuss some potential extensions and issues which may affect the interpretation of our results.

A. *Finer Earnings Ranges*

The analysis above was predicated on the coarsening of earnings dictated in (7). This coarsening scheme is "natural" in the sense that the JF reform changed the utility of all the alternatives corresponding to each of the earning ranges in (7) in the *same* direction. Nevertheless, it can be of interest to consider finer coarsenings of earnings. For instance, our finding in Table 5 of a significant opt-in response could hypothetically reflect trivial earnings reductions from \$1 above the poverty line to exactly the poverty line. To assess such possibilities, consider the following finer coarsening of earnings obtained by partitioning range 2 into two subranges:

$$\tilde{E}_i \equiv \begin{cases} 0 & \text{if } E = 0 \\ 1 & \text{if } E \leq FPL_i \\ 2' & \text{if } E \in (FPL_i, 1.2 \times FPL_i] \\ 2'' & \text{if } E > 1.2 \times FPL_i \end{cases}. \tag{16}$$

In the online Appendix, we derive bounds on the response probabilities $\pi_{2'n,1r}$ and $\pi_{2''n,1r}$ which correspond to the fraction of women in earnings ranges $2'$ and $2''$ who opt in to assistance by reducing their earnings. These bounds exploit the fact that, by revealed preference, no woman will pair state $2'u$ under AFDC with state $2''u$ under JF. Likewise, no woman will pair state $2'n$ under AFDC with state $2''n$ under JF.

Implementing these formulas, we find that at least 27 percent of women who would work off assistance in earnings range $2''$ under AFDC reduced their earnings below the poverty line in response to the JF reform. Accounting for sampling uncertainty yields a lower limit on a 95 percent confidence interval for $\pi_{2''n,1r}$ of 0.17. Hence, the evidence is strong that some large earnings reductions occurred in response to the JF reform. We also find that at least 31 percent of women who would

[25] Although the expressions for the bounds differ depending on whether the utility function obeys (4) or (5), the solutions that bind in the data are the same. This is because inequality (13) holds in the sample.

work off assistance in earnings range $2'$ under AFDC reduced their earnings below the poverty line in response to reform. The lower limit of the confidence interval for $\pi_{2'n,1r}$ is 0.19, which indicates that some opt-in responses also took place from earnings ranges closer to the poverty line.[26]

It is also possible to partition range 1 into two or more subranges. We have experimented with such extensions but found that they fail to offer additional insights regarding the effects of the JF reform.[27]

B. *A Test for Anticipation*

As mentioned in Section IV, the JF time limits create incentives for a risk-averse woman to save months of welfare eligibility for later periods when her earnings may be lower (e.g., due to job loss). Thus, under some conditions, JF may actually make working on welfare less attractive, as this choice requires sacrificing the option value of using welfare an additional month in the future. This could, in turn, lead to a violation of our model's static revealed preference restrictions.

Following Grogger and Michalopoulos (2003), we conduct a simple test for whether the JF time limits in fact yielded anticipatory effects. Our test compares the impact of reform on the welfare use of women who at baseline had a youngest child 16–17 years old (for whom the time limits were irrelevant) to impacts on the welfare use of women who had younger children. As shown in Table 6, we cannot reject the null hypothesis that the average impact of JF on monthly welfare take-up is the same for both groups of women. In fact, our point estimates suggest that the response of women with younger children to reform was actually slightly greater than the response of women with children 16–17 years old, which is the opposite of what anticipatory behavior would suggest. While this finding does not prove that the women in our sample were myopic, it does suggest that anticipatory responses to the time limits were probably small.[28]

[26] Note that some of the responses involving reductions from earnings range $2'$ to range 1 could be larger than those from earnings range $2''$ to range 1 since we don't know which earnings level in range 1 is being selected. The upper bounds on these response probabilities are uninformative.

[27] There are good reasons for this. Recall that the theory does not constrain the sign of the labor supply responses that occur within range 1. This theoretical indeterminacy persists if range 1 is partitioned into subranges and prevents identification of the magnitude of these allowed intensive margin responses. Additionally, the possibility of underreporting limits the utility of revealed preference arguments because states $1u$ and $1r$ are not empirically distinguishable. This prevents identification of the magnitude of any responses to the reform entailing adjustments in reporting behavior within range 1.

[28] Grogger and Michalopoulos (2003) rely on data from a randomized welfare reform where the experimental group was exposed to a 24-month time limit (or a 36-month limit if particularly disadvantaged). JF's more stringent 21-month time limit might be expected to produce a larger anticipatory response than found by Grogger and Michalopoulos. It does not. One possible explanation for this discrepancy is that, as mentioned in Section I, a large fraction of JF experimental units were exempted from time limits, and a large fraction of the nonexempted units were granted six-month extensions. Bloom et al. (2002, p. 59) report that "written material produced by the DSS explicitly stated that extensions would be possible." Also, "staff reported that many recipients were initially skeptical that the time limit would be implemented (in fact, many staff said that they themselves were skeptical)." Based on the Interim Client Survey, it appears that "from the beginning, most recipients understood that the time limit would not necessarily result in cancellation of their welfare grant."

TABLE 6—FRACTION OF MONTHS ON WELFARE BY EXPERIMENTAL STATUS AND AGE OF YOUNGEST CHILD

Age of youngest child at baseline	16 or 17		15 or less
AFDC	0.185		0.383
	(0.078)		(0.069)
JF	0.251		0.473
	(0.077)		(0.069)
Difference	0.066		0.090
	(0.054)		(0.010)
Difference-in-differences		−0.024	
		(0.054)	

Notes: Sample consists of 87,717 case-months: 21 months of data on each of 4,177 cases with nonmissing baseline information on age of youngest child. Table gives regression-adjusted fraction of case-months that women participated in welfare by experimental status and age of youngest child at baseline. Robust standard errors computed using clustering at case level.

XI. Conclusion

Our analysis of the Jobs First experiment suggests that women responded to the policy incentives of welfare reform along several margins, some of which entail an intensive margin and some of which entail an extensive margin labor supply response. This finding is in accord with BGH's original interpretation of the JF experiment and with recent evidence from Blundell, Bozio, and Laroque (2011a,b), who find that secular trends in aggregate hours worked appear to be driven by both intensive and extensive margin adjustments. Our conclusions are also qualitatively consistent with recent studies relying on dynamic parametrically structured labor supply models (e.g., Blundell, Pistaferri, and Saporta-Eksten 2016; Blundell et al. 2013).

An important question is the extent to which our finding of intensive margin responsiveness might generalize to other transfer programs that lack sharp budget notches but still involve phase-out regions that should discourage work. It seems plausible that the JF notch would yield larger disincentive effects than, say, the budget kink induced by the EITC phase-out region. However, Bitler, Gelbach, and Hoynes (2008) show that experimental responses to a Canadian reform inducing such a gradual benefit phase-out generated a pattern of earnings QTEs similar to that found in the JF experiment. More conclusive evidence on this question may be had via an application of the methods developed here to other policy reforms.

Though we studied a randomized experiment, our approach is easily adapted to quasi-experimental settings. Estimates of the relevant counterfactual choice probabilities can be formed using one's research design of choice (e.g., a difference-in-differences design), subject to the usual caveat that different designs may identify counterfactuals for different treated subpopulations.[29] With the two sets of marginal choice probabilities, bounds on response probabilities can then be had by a direct application of the methods developed in this paper.

[29] For example, if one uses an instrumental variables design, counterfactuals are, under weak assumptions, identified only for the subpopulation of "compliers" (Imbens and Rubin 1997).

A potentially fruitful avenue for future research is to consider the application of revealed preference arguments to dynamic models. Alternatives in such models consist of sequences of possible choices, which significantly enlarges the space of potential responses that can occur. However, explicitly dynamic models also provide additional opportunities to incorporate plausible nonparametric restrictions (e.g., stationary and time-separable preferences) that may yield interesting empirical predictions.

Finally, it is worth noting that decision makers have sometimes been found to violate revealed preference restrictions empirically (e.g., Choi et al. 2014). An interesting question for future research is whether comparably informative nonparametric restrictions can be derived from behavioral models designed to explain systematic deviations from rationality.

REFERENCES

Altonji, Joseph G., and Christina H. Paxson. 1992. "Labor Supply, Hours Constraints, and Job Mobility." *Journal of Human Resources* 27 (2): 256–78.

Andrews, Donald W. K., and Panle Jia Barwick. 2012. "Inference for Parameters Defined by Moment Inequalities: A Recommended Moment Selection Procedure." *Econometrica* 80 (6): 2805–26.

Andrews, Donald W. K., and Sukjin Han. 2009. "Invalidity of the Bootstrap and the *m* out of *n* Bootstrap for Confidence Interval Endpoints Defined by Moment Inequalities." *Econometrics Journal* 12 (1): S172–99.

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

Barrett, Garry F., and Stephen G. Donald. 2003. "Consistent Tests for Stochastic Dominance." *Econometrica* 71 (1): 71–104.

Beffy, Magali, Richard Blundell, Antoine Bozio, and Guy Laroque. 2014. "Labour Supply and Taxation with Restricted Choices." Institute for Fiscal Studies Working Paper W14/04.

Bitler, Marianne P., Jonah B. Gelbach, and Hilary W. Hoynes. 2005. "Distributional Impacts of the Self-Sufficiency Project." National Bureau of Economic Research Working Paper 11626.

Bitler, Marianne P., Jonah B. Gelbach, and Hilary W. Hoynes. 2006. "What Mean Impacts Miss: Distributional Effects of Welfare Reform Experiments." *American Economic Review* 96 (4): 988–1012.

Bitler, Marianne P., Jonah B. Gelbach, and Hilary W. Hoynes. 2008. "Distributional Impacts of the Self-Sufficiency Project." *Journal of Public Economics* 92 (3-4): 748–65.

Bitler, Marianne P., Jonah B. Gelbach, and Hilary W. Hoynes. 2014. "Can Variation in Subgroups' Average Treatment Effects Explain Treatment Effect Heterogeneity? Evidence from a Social Experiment." National Bureau of Economic Research Working Paper 20142.

Blomquist, Soren, Anil Kumar, Che-Yuan Liang, and Whitney K. Newey. 2014. "Individual Heterogeneity, Nonlinear Budget Sets, and Taxable Income." Institute for Fiscal Studies, Centre for Microdata Methods and Practice Working Paper CWP21/14.

Bloom, Dan, Susan Scrivener, Charles Michalopoulos, Pamela Morris, Richard Hendra, Diana Adams-Ciardullo, Johanna Walter, and Wanda Vargas. 2002. *Jobs First: Final Report on Connecticut's Welfare Reform Initiative.* New York: Manpower Demonstration Research Corporation.

Blundell, Richard, Antoine Bozio, and Guy Laroque. 2011a. "Extensive and Intensive Margins of Labour Supply: Working Hours in the US, UK and France." IZA Discussion Paper 6051.

Blundell, Richard, Antoine Bozio, and Guy Laroque. 2011b. "Labor Supply and the Extensive Margin." *American Economic Review* 101 (3): 482–86.

Blundell, Richard, Monica Costas Dias, Costas Meghir, and Jonathan Shaw. 2013. "Female Labour Supply, Human Capital and Welfare Reform." National Bureau of Economic Research Working Paper 19007.

Blundell, Richard, Luigi Pistaferri, and Itay Saporta-Eksten. 2016. "Consumption Inequality and Family Labor Supply." *American Economic Review* 106 (2): 387–435.

Burtless, Gary, and Jerry A. Hausman. 1978. "The Effect of Taxation on Labor Supply: Evaluating the Gary Negative Income Tax Experiments." *Journal of Political Economy* 86 (6): 1103–30.

Card, David, and Dean R. Hyslop. 2005. "Estimating the Effects of a Time-Limited Earnings Subsidy for Welfare-Leavers." *Econometrica* 73 (6): 1723–70.

Chan, Marc. K. 2013. "A Dynamic Model of Welfare Reform." *Econometrica* 81 (3): 941–1001.

Chernozhukov, Victor, Sokbae Lee, and Adam M. Rosen. 2013. "Intersection Bounds: Estimation and Inference." *Econometrica* 81 (2): 667–737.

Chetty, Raj. 2012. "Bounds on Elasticities with Optimization Frictions: A Synthesis of Micro and Macro Evidence on Labor Supply." *Econometrica* 80 (3): 969–1018.

Chetty, Raj, John N. Friedman, Tore Olsen, and Luigi Pistaferri. 2011. "Adjustment Costs, Firm Responses, and Micro vs. Macro Labor Supply Elasticities: Evidence from Danish Tax Records." *Quarterly Journal of Economics* 126 (2): 749–804.

Chetty, Raj, Adam Guren, Day Manoli, and Andrea Weber. 2011. "Are Micro and Macro Labor Supply Elasticities Consistent? A Review of Evidence on the Intensive and Extensive Margins." *American Economic Review* 101 (3): 471–75.

Choi, Syngjoo, Shachar Kariv, Wieland Mueller, and Dan Silverman. 2014. "Who Is (More) Rational?" *American Economic Review* 104 (6): 1518–50.

Cogan, John F. 1981. "Fixed Costs and Labor Supply." *Econometrica* 49 (4): 945–63.

Couch, Kenneth A. 2014. "Experimental Evidence of Welfare Reform on Employment and Earnings." Unpublished.

Diamond, P. 1980. "Income Taxation with Fixed Hours of Work." *Journal of Public Economics* 13 (1): 101–10.

Dickens, William T., and Shelly J. Lundberg. 1993. "Hours Restrictions and Labor Supply." *International Economic Review* 34 (1): 169–92.

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

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

Hansen, Gary D. 1985. "Indivisible Labor and the Business Cycle." *Journal of Monetary Economics* 16 (3): 309–27.

Heckman, James J. 1983. "Comment." In *Behavioral Simulation Methods in Tax Policy Analysis*, edited by Martin Feldstein, 70–82. Chicago: University of Chicago Press.

Heckman, James J. 1993. "What Has Been Learned about Labor Supply in the Past Twenty Years?" *American Economic Review* 83 (2): 116–21.

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

Hotz, V. Joseph, Charles H. Mullin, and John Karl Scholz. 2003. "The Effects of Welfare Reform on Employment and Income: Evidence from California." Unpublished.

Hoynes, Hilary Williamson. 1996. "Welfare Transfers in Two-Parent Families: Labor Supply and Welfare Participation under AFDC-UP." *Econometrica* 64 (2): 295–332.

Imbens, Guido W., and Charles F. Manski. 2004. "Confidence Intervals for Partially Identified Parameters." *Econometrica* 72 (6): 1845–57.

Imbens, Guido W., and Donald B. Rubin. 1997. "Estimating Outcome Distributions for Compliers in Instrumental Variables Models." *Review of Economic Studies* 64 (4): 555–74.

Keane, Michael, and Robert Moffitt. 1998. "A Structural Model of Multiple Welfare Program Participation and Labor Supply." *International Economic Review* 39 (3): 553–89.

Keane, Michael P., and Kenneth I. Wolpin. 2002. "Estimating Welfare Effects Consistent with Forward-Looking Behavior: Part II: Empirical Results." *Journal of Human Resources* 37 (3): 600–622.

Keane, Michael P., and Kenneth I. Wolpin. 2007. "Exploring the Usefulness of a Nonrandom Holdout Sample for Model Validation: Welfare Effects on Female Behavior." *International Economic Review* 48 (4): 1351–78.

Keane, Michael P., and Kenneth I. Wolpin. 2010. "The Role of Labor and Marriage Markets, Preference Heterogeneity, and the Welfare System in the Life Cycle Decisions of Black, Hispanic, and White Women." *International Economic Review* 51 (3): 851–92.

Kitamura, Yuichi, and Jorg Stoye. 2013. "Nonparametric Analysis of Random Utility Models: Testing." Cowles Foundation for Research in Economics, Yale University, Cowles Foundation Discussion Paper 1902.

Kleven, Henrik J., and Mazhar Waseem. 2013. "Using Notches to Uncover Optimization Frictions and Structural Elasticities: Theory and Evidence from Pakistan." *Quarterly Journal of Economics* 128 (2): 669–723.

Kline, Patrick, and Melissa Tartari. 2016. "Bounding the Labor Supply Responses to a Randomized Welfare Experiment: A Revealed Preference Approach: Dataset." *American Economic Review.* http://dx.doi.org/10.1257/aer.20130824.

Laroque, Guy. 2005. "Income Maintenance and Labor Force Participation." *Econometrica* 73 (2): 341–76.

Ljungqvist, Lars, and Thomas J. Sargent. 2011. "A Labor Supply Elasticity Accord?" *American Economic Review* 101 (3): 487–91.
MaCurdy, Thomas, David Green, and Harry J. Paarsch. 1990. "Assessing Empirical Approaches for Analyzing Taxes and Labor Supply." *Journal of Human Resources* 25 (3): 415–90.
Manski, Charles F. 2014. "Identification of Income-Leisure Preferences and Evaluation of Income Tax Policy." *Quantitative Economics* 5 (1): 145–74.
McFadden, Daniel L. 2005. "Revealed Stochastic Preference: A Synthesis." *Economic Theory* 26 (2): 245–64.
Meyer, Bruce D. 2002. "Labor Supply at the Extensive and Intensive Margins: The EITC, Welfare, and Hours Worked." *American Economic Review* 92 (2): 373–79.
Meyer, Bruce D., and Dan 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, Robert. 1983. "An Economic Model of Welfare Stigma." *American Economic Review* 73 (5): 1023–35.
Moffitt, Robert. 1990. "The Econometrics of Kinked Budget Constraints." *Journal of Economic Perspectives* 4 (2): 119–39.
Rogerson, Richard. 1988. "Indivisible Labor, Lotteries and Equilibrium." *Journal of Monetary Economics* 21 (1): 3–16.
Roy, A. D. 1951. "Some Thoughts on the Distribution of Earnings." *Oxford Economic Papers* 3: 135–46.
Saez, Emmanuel. 2002. "Optimal Income Transfer Programs: Intensive versus Extensive Labor Supply Responses." *Quarterly Journal of Economics* 117 (3): 1039–73.
Saez, Emmanuel. 2010. "Do Taxpayers Bunch at Kink Points?" *American Economic Journal: Economic Policy* 2 (3): 180–212.
Slemrod, Joel. 2010. "Buenas Notches: Lines and Notches in Tax System Design." Unpublished.
Swann, Christopher A. 2005. "Welfare Reform When Recipients Are Forward-Looking." *Journal of Human Resources* 40 (1): 31–56.

This article has been cited by:

1. 2016. Book Reviews. *Journal of Economic Literature* **54**:3, 945-946. [Abstract] [View PDF article] [PDF with links]
2. 2016. Book Reviews. *Journal of Economic Literature* **54**:3, 942-958. [Citation] [View PDF article] [PDF with links]
3. 2016. Book Reviews. *Journal of Economic Literature* **54**:3, 944-945. [Abstract] [View PDF article] [PDF with links]