What Distributional Impacts Mean: Welfare Reform Experiments and Competing Margins of Adjustment

Patrick Kline — UC Berkeley/NBER

Melissa Tartari* — Yale

February, 2013

Abstract

We study the impact of the Connecticut Jobs First (JF) welfare reform experiment on the labor supply and program participation decisions of a sample of welfare applicants and recipients. A rich optimizing model is developed incorporating underreporting decisions, labor supply constraints, fixed costs of work, and welfare stigma and hassle. Qualitatively, our model rationalizes the large empirical impacts of the JF experiment on the distribution of earnings despite the absence of bunching at a program induced notch in agents' budget sets. We show that the model places nonparametric restrictions on experimental impacts that can be used to develop bounds on the magnitude of a variety of intensive and extensive margin responses to reform. Our results indicate that, in addition to incentivizing work at the extensive margin, the JF experiment induced a substantial welfare "opt-in" response among women with relatively high earnings potential.

*We thanks Andres Santos and seminar participants at Bates-White, the New Economic School, Queens University, UC Berkeley, and Yale for useful comments. This research was supported by NSF Grant #0962352.

A large empirical literature studies behavioral responses to transfer policies (Moffitt, 2002 provides a review). Traditional neoclassical models of labor supply predict that the price distortions induced by such policies will yield complex intensive margin labor supply responses, with some households working more and some less, often with discrete bunching of earnings at particular levels (Hausman, 1985; Moffitt, 1990). However, many studies find that key qualitative predictions of these frictionless models fail empirically. In both survey and administrative data, earnings tend not to exhibit much bunching at kinks in agents' budget sets (Heckman, 1983; Saez, 2010). Moreover, several studies exploiting policy variation fail to find evidence of important intensive margin responses (Ashenfelter, 1983; Eissa and Liebman, 1996; Eissa and Hoynes, 2006; Meyer and Rosenbaum, 2001; Meyer, 2002). These facts have led many to conclude that behavioral responses on the intensive margin are often sharply constrained (e.g., as in Chetty et al., 2011) and that adjustment to policy reforms is likely to occur primarily on the extensive margin.[1]

In an important contribution, Bitler, Gelbach, and Hoynes (BGH, 2006) provide non-parametric evidence that intensive margin responses may have played a substantial role in mediating the effects of welfare reform. They show, using experimental data from Connecticut's Jobs First (JF) program, that welfare reform induced a nuanced pattern of quantile treatment effects (QTEs) on earnings consistent with the unconstrained neoclassical labor supply model. In particular, they find that the Jobs First program boosted the middle quantiles of earnings while lowering the top quantiles, yielding a mean earnings effect near zero. BGH argue that the negative impacts on upper quantiles provide suggestive evidence of an "opt-in" response to welfare (Ashenfelter, 1983), whereby some workers who would have worked without welfare in the absence of reform are induced to lower their earnings in order to qualify for benefits. The finding of a strong opt-in effect suggests important intensive margin responsiveness among the population of poor women with children, setting this study apart from most prior evidence in the literature. Such an interpretation, if correct, would appear to warrant a rethinking of the optimal design of transfer policy (Saez, 2002; Laroque, 2005).

In this paper, we extend BGH's analysis of the Jobs First experiment to formally identify the magnitude of intensive and extensive margin adjustments to the program's various work

[1] Heckman (1993), for instance, concludes that "elasticities are closer to 0 than 1 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 "indivisibile" (Hansen, 1985; Rogerson, 1988; Ljungqvist and Sargent, 2011).

incentives. To conclude that adjustment occurred along the intensive margin, one must infer features of the *joint* distribution of women's potential earnings under the two policy regimes. As noted by Heckman, Smith, and Clements (1997), joint distributions are only identified by QTEs under the "rank invariance" assumption that potential earnings under the two regimes exhibit perfect positive dependence,[2] a condition which BGH acknowledge is likely to be violated in their sample.[3] Our analysis replaces the rank invariance assumption with a set of joint restrictions on earnings and program participation behavior derived from an optimizing model where agents face constraints on their earnings opportunities. We show that these restrictions are refutable with experimental data and allow the development of informative bounds on the probability of opting into welfare along with the probability of responding along a number of other margins, including the under-reporting of earnings to case workers. We develop estimators and inferential methods for these response probabilities and find that the Jobs First program induced a large opt-in response along with a substantial extensive margin response. To our knowledge, these estimates are the first to separately identify the extensive and intensive margin responses to a policy reform without parametric assumptions.

We begin our analysis by developing the standard frictionless model of labor supply and deriving its key qualitative predictions for the impact of the JF program on earnings and program participation choices. Empirically, most of these predictions appear to be satisfied in the data. We show that the program induced many women to work, raised welfare participation, and increased the incidence of work while on welfare. We also show that the program "warped" the distribution of earnings in the theoretically expected manner, with work becoming more common at low earnings levels and less common at high earnings levels. Finally, we show that the pattern of distributional impacts varies across subgroups in a manner consistent with theory.

However, one notable prediction of the model does not hold. The opt-in effects of welfare reform should be accomplished, in part, via a point mass of women locating at the JF eligibility notch. Using administrative data sources, we show that there is no evidence of bunching at this notch, despite its obvious salience. We suggest this is likely due to a combination of constraints on workers' ability to adjust their earnings levels and under-reporting behavior. Indeed, many women with ineligible earning levels nonetheless receive benefits under JF, suggesting that under-reporting of earnings is likely to be common among

[2]Potential outcomes would also be identified under an assumption of perfect *negative* dependence of potential outcomes, which is implausible in most settings.

[3]In a related analysis, BGH (2005) find strong evidence of rank reversals in the Canadian Self-Sufficiency Project experiment.

program participants, a finding in line with much of the previous literature (Greenberg, Moffitt, and Friedman, 1981; Greenberg and Halsey, 1983; Hotz, Mullin, and Scholz, 2003).

We then turn to developing the quantitative implications of an augmented model where workers may lack fine control over their earnings but are able to make crude intensive margin adjustments, to switch between program participation states, and to under-report their earnings. The addition of under-reporting behavior introduces additional margins of adjustment that have not typically been incorporated into prior models but are potentially quite important for a number of transfer programs. Notably, the addition of under-reporting behavior yields violations of the rank invariance condition as reform may induce some women to earn relatively large amounts which are then under-reported to case workers. Unlike traditional parametric models of nonlinear budget sets (e.g. Burtless and Hausman, 1978; Hoynes, 1996; Keane and Moffitt, 1998), our constrained model has no refutable predictions for the cross-sectional distribution of earnings and program participation under a given policy regime.[4] We show however that the model places strong restrictions on the set of potential transitions between coarse earnings and welfare participation categories ("response margins") that may be induced by the experiment. We use these theoretical restrictions to derive analytical bounds on the probabilities of making transitions along each allowable margin.

Applying our identification results, we find evidence of substantial intensive and extensive margin responses to treatment. Jobs First appears to have incentivized many women who would not have worked to do so and many who would have worked off welfare at low earnings to participate in the JF program. Corroborating the analysis of BGH, we find a strong opt-in response among women who would have worked at relatively high earnings levels. We also find evidence that the JF work requirements induced some women to work but under-report their earnings in order to maintain eligibility for benefits.

Our results demonstrate that a theoretically motivated coarsening of earnings data can allow researchers to bound the size of competing response margins under weaker assumptions than have been previously considered in the literature. This insight appears to be new and may be applicable to other settings with continuous choices (e.g. capital taxation or electricity demand) where incentives are highly nonlinear. More generally, using theoretical restrictions on the set of potential transitions induced by reform provides a feasible approach to going beyond mean effects and estimating impact distributions, a goal which has been argued by many to be important for appropriately interpreting experimental impacts

[4] See Macurdy, Green, and Paarsch (1990) for an early critique of parametrically structured econometric models of labor supply with nonlinear budget sets.

(Heckman and Smith, 1995; Heckman, Smith, and Clements, 1997; Heckman, 2001; Deaton, 2009).

The bounding approach developed here is similar in spirit to the recent work of Manski (2012) who uses revealed preference arguments to set-identify features of policy counterfactuals. However, our analysis explores the identifying power of experimental variation in isolating response margins, a subject he does not consider. We also work with a richer model of labor supply behavior incorporating fixed costs of work, under-reporting behavior, the effects of welfare hassle, and constraints on earning opportunities. Blundell, Bozio, and Laroque (2011a,b) also have a recent 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. They find that adjustments along both the intensive and extensive margins are important contributors to fluctuations in aggregate hours worked.

The remainder of the paper is structured as follows: Section 1 describes the Jobs First Experiment while Section 2 lays out a simplified version of our model without constraints or under-reporting behavior. Section 3 describes the data used and Section 4 provides experimental impacts on a variety of outcomes and assesses their consistency with the baseline model's predictions. Section 5 documents the absence of bunching of earnings at the eligibility notch using administrative data. Section 6 describes our full model incorporating earnings constraints and under-reporting and Section 7 studies identification of response margins under the model's restrictions. Section 8 provides our empirical results regarding response probabilities and Section 9 concludes.

1 The Jobs First Program

With the passage of the Personal Responsibility and Work Opportunity Reconciliation Act (PRWORA) in 1996, all fifty states were required to replace their Aid to Families with Dependent Children (AFDC) programs with Temporary Assistance to Needy Families (TANF) programs. This change involved a series of major reforms including the imposition of time limits, work requirements, and improved financial incentives to work. The state of Connecticut responded to PRWORA by implementing the Jobs First (JF) program which involved each of the major features of the TANF program including a particularly strong (and salient) change in the financial incentives to work while on assistance.

To study the effectiveness of the reform, the state contracted with the Manpower Development Research Corporation (MDRC) to conduct a randomized evaluation, comparing the Jobs First TANF program to the earlier state AFDC program. Between January 1996 and February 1997, MDRC collected a baseline sample of about 4,800 single parent AFDC recipients and applicants and randomly assigned them to either the new JF program or the old AFDC program with equal probability. To conduct the evaluation, administrative data on earnings and welfare participation were collected for several years prior to and following the date of random assignment and merged with a baseline survey conducted by MDRC. This experiment has been heavily studied, with several analyses finding that the program encouraged work and had important effects on the distribution of both earned and unearned income (Adams-Ciardullo et al., 2002; BGH, 2006, 2010).

Table 1 provides a summary of the key JF and AFDC program features. A distinguishing feature of the JF program is the 100% earnings disregard which provides a dramatic reduction in the implicit tax on earnings faced by women on welfare relative to AFDC. The JF earnings disregard was meant to incentivize work at low earnings levels. In doing so it created an eligibility "notch" in the transfer scheme, with a windfall loss of the entire grant amount occurring if a woman earns a dollar more than the federal poverty line. This notch creates strong incentives not to earn amounts just above the poverty line and, when possible, to under-report income in order to maintain program eligibility.[5]

Another important feature of JF, like all TANF programs, is the imposition of time limits. As documented in MDRC's final report (Adams-Ciardullo et al., 2002) and BGH (2006), this feature induced time varying effects of the program as JF cases eventually became ineligible for program benefits. For this reason, we restrict our analysis to the first 7 quarters of the experiment, a horizon over which time no case was in danger of reaching the limit.[6] Finally, the JF program involved a series of formal work requirements and imposed sanctions on cases that failed to seek work. Evidence from the final report suggests the sanctions were not heavily enforced. Nevertheless, in our empirical work, we allow for the possibility that the JF work requirements made life less pleasant for women who chose to stay on welfare without working.

[5] The program also induced a second notch for the sample of JF *applicants* who faced a strict earnings test in order to *establish* eligibility. AFDC did not have an earnings test, but benefits for that program phased out at an amount above the JF earnings test. Hence, it became harder under JF for high earning applicants to establish eligibility. Because our analysis is static, we follow BGH in ignoring this feature of the policy. See Card and Hyslop (2005) and Ferrall (2012) for empirical analyses of dynamic responses to eligibility incentives. We have experimented with restricting our analysis to recipients only and found similar results however this group contains substantially fewer high earners which limits our ability to detect opt-in behavior.

[6] If agents are forward looking, this restriction may not fully remove the influence of the time limits on behavior (Grogger and Michalopoulos, 2003). BGH (2006) argue that this is probably not a major concern in this sample but we cannot fully rule out the possibility that our results would be different if time limits were not present.

2 Budget Set and Optimizing Behavior

Figure 1 provides a stylized depiction of the budget faced by a woman with 2 children who is new to welfare under the AFDC and JF policy rules respectively. The vertical axis of the graph gives total income (earned income + transfers) while the horizontal axis gives earned income E. The JF budget set contains a large notch at the federal poverty line beyond which welfare eligibility expires. Under JF, total income received when monthly earnings equal the federal poverty line (FPL) exceeds that for earnings in the range $(FPL, FPL+\overline{G})$ where $\overline{G}$ is the base grant amount which is common to JF and AFDC. This is to be contrasted with the AFDC budget set which phases out smoothly, yielding an implicit tax rate of approximately 49% on earnings above a \$120 disregard.

We summarize these policy rules with 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 may vary across women with the same earnings due to variation in the size of the assistance unit (AU). Thus,

$$\begin{aligned} G_i^a(E) &= \max\left\{\overline{G}_i - 1\left[E > \delta_i\right](E - \delta_i)\tau_i, 0\right\} \\ G_i^j(E) &= 1\left[E < FPL_i\right]\overline{G}_i \end{aligned} \tag{1}$$

where δ_i and $1 - \tau_i$ are the fixed and proportional AFDC earnings disregards, both of which may vary across households depending upon how long the woman has been receiving benefits.[7] Note that the base grant amount $\overline{G}_i$ and the federal poverty line FPL_i also vary across households due to differences in AU size.

To study the effects of the program, we begin by considering a conventional static utility maximization framework where agents have complete control over their earnings. We assume women in our sample have heterogenous convex non-satiated utility functions

$$U_i(E, C)$$

defined over earnings and a consumption equivalent C. As in Saez (2010), utility is increasing in consumption C but decreasing in earnings E which require effort to generate. In this framework, one can think of heterogeneity in the disutility of earnings as worker skill. The consumption equivalent C incorporates a variety of psychic and monetary costs and takes

[7] Under AFDC, women are allowed four months use of the proportional disregard before it expires.

the form:

$$C = E + \left(G_i^t(E) - \phi_i - \eta_i^t 1[E = 0]\right) D - \mu_i 1[E > 0]. \tag{2}$$

The variable $D \in \{0, 1\}$ is an indicator for the woman participating in welfare. The parameter $\phi_i \geq 0$ gives the dollar value of welfare stigma which may vary across women and explain why some women with no earnings don't participate in welfare. We also allow for a fixed cost of work μ_i which may vary across women. Fixed costs discourage work at low earnings levels.[8] Finally, $\eta_i^t \geq 0$ gives the dollar value of the hassle a woman faces from authorities when not working and receiving benefits. This hassle may depend on the policy regime. Because JF includes stronger work requirements and sanctions, we assume that

$$\eta_i^j \geq \eta_i^a \ \forall i.$$

We leave the joint distribution of $\left(\phi_i, \eta_i^a, \eta_i^j, \mu_i, U_i(.)\right)$ unrestricted.

The woman's decision problem is to:

$$\max_{E \geq 0,\ D \in \{0,1\}} U_i(E, C) \quad \text{subject to } (1) \text{ and } (2)$$

We now summarize our predictions from this model. Consider first a woman at point A in Figure 1 who, under AFDC, would participate in welfare without working. She may be incentivized by JF rules to work while on welfare both by the reduction in implicit tax rates on earnings and the hassle associated with JF. If fixed costs are large enough, the additional hassle associated with not working under JF may induce her to leave welfare entirely and earn more than the federal poverty line. Alternatively, she may simply respond to the hassle by opting out of welfare and continuing to not work. Finally, she may not respond at all and simply continue to stay on welfare and not work.

A woman working on welfare at point B under AFDC 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 any case will lead her to continue working on welfare. If, as much of the literature on female labor supply suggests (Blundell and Macurdy, 1999), the substitution effect dominates the income effect, a woman will be expected to work more but not more than the federal poverty line as this level of earnings was in her choice set under AFDC and not chosen.

A woman who chooses to locate at point A' under AFDC must face high welfare stigma as she is forgoing the full grant amount $\overline{G}_i$. Not working under JF will be at least as

[8]See Meyer and Heim (2004) for a careful analysis of labor supply behavior with fixed costs of work.

unattractive since the base grant amount is unaffected by the reform. She will also not want to work on JF because the benefits of work are no greater under JF than off welfare and she chose not to work while off welfare. Therefore, her behavior will be unaffected by reform.

A woman who works off welfare at point B' under AFDC must also dislike welfare participation. Yet she may be induced by the reforms to work under JF. This is likely if stigma is strong but so is the substitution effect. In such a case, the woman will value the reduction in tax rate under JF enough to justify participating in welfare.

A woman working on welfare under AFDC at point C in Figure 1 will choose to participate in JF which would offer an increase in income for less work with the same amount of stigma. Indeed, the windfall associated with JF participation may well induce an income effect leading her to reduce her earnings below the federal poverty line. Much the same response is likely for a woman who would work off welfare under AFDC at point D. If however the woman faces severe stigma with regard to welfare participation she may nonetheless choose to remain off welfare and continue to earn at point D.

A high earning woman at point E in Figure 1 may face incentives to opt-in to welfare under JF which would involve lowering her earnings to the federal poverty line.[9] At this point she will unambiguously want to work more as participation involves a reduction in income for her. She will not do so because of the large eligibility notch she faces. Thus, we have the sharp nonparametric prediction that opt-in should lead to a mass point of workers locating exactly at the poverty line.

The key predictions of the model then, are that:

- The fraction of women who work should increase under JF
- The fraction of women on welfare and not working should decrease
- A mass of workers should locate exactly at the FPL under JF

We turn now to an examination of whether these predictions are actually satisfied in the data.

[9] Note that women who would locate at points C or D under AFDC may also choose to locate exactly at the poverty line under JF.

3 Data

Our data come from the MDRC Jobs First Public Use Files. The data contain a baseline survey of demographic and family composition variables merged with administrative information on welfare participation, rounded welfare payments, family composition, and rounded earnings covered by the state unemployment insurance (UI) system. There are a number of important limitations to the Public Use Files that place constraints on our analysis. While welfare payments are measured monthly, UI earnings data are only available quarterly. Because we are studying the relationship between program participation and earnings decisions, it is important that we be able to compare these variables on a consistent time scale. The Public Use Files do not directly provide the month of randomization, making it difficult to aggregate the monthly welfare information to the quarterly level. We infer the month of randomization by contrasting monthly assistance payments with an MDRC constructed variable providing quarterly assistance payments. For each case, we have found that a unique month of randomization leads the aggregation of the monthly payments to match the quarterly measure to within rounding error.

Another difficulty is that our administrative measure of the size of the assistance unit is missing for most cases. For the Jobs First sample we are able to infer an AU size in most months from the grant amount while the women are on welfare. If however AU size changes while off welfare we are not able to detect this change. Moreover, in some cases the grant amount is reduced, presumably because the household reported some unearned income. In both cases we use the grant amount in other months to impute AU size in a month for which we cannot observe it. For the AFDC sample, the grant amount depends on many unobserved factors, preventing us from inferring the AU size even while on welfare.

In cases where we need to compute treatment effects by AU size we rely on another variable (kidcount) collected in the baseline survey which asks for the number of children in the household at the time of random assignment. This variable is top-coded at three children. Appendix Table 1 gives a cross-tabulation in the JF sample of kidcount with our more reliable AU size measure inferred from grant amounts. The tabulation suggests the kidcount variable is a reasonably accurate measure of AU size over the first 7 quarters post-random assignment conditional on the number of children at baseline being less than three. As might be expected, the kidcount variable tends to underestimate the true AU size as women may have additional children over the 7 quarters following the baseline survey. To deal with this problem we try inflating the kidcount based AU size by one in order to avoid understating the location of the poverty line for most households.

Table 2 shows descriptive statistics for our analysis sample. We have 4,802 cases with complete earnings and pre-random assignment characteristics, yielding essentially the same baseline sample as in BGH (2006).[10] As the Table makes clear, 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 (2006) in using propensity score methods to adjust for these baseline differences in order to increase the efficiency of our estimation procedures. These techniques are described in the Appendix. After adjustment, the means of the AFDC and JF groups are very similar.

We also examine three subgroups of the data characterized by their earnings seven quarters prior to random assignment. These groups were studied previously by BGH (2010), with the zero earnings group having no earnings in pre-assignment quarter 7, and the low and high earning groups having quarterly earnings below or above the median of the non-zero observations respectively. Descriptive statistics for these groups are also provided in Table 2. Because pre-assignment earnings proxy for tastes and earnings ability, the switch to JF likely presented these groups with different incentives. We document substantial treatment effect heterogeneity between them later on.

The Eligibility Notch and Under-reporting Behavior

Before examining the effects of JF it is useful to verify that transfers do in fact obey the JF policy rules. Figure 2 plots the empirical relationship between quarterly UI earnings and welfare payments for women in the JF sample. Because the poverty line varies by AU size we rescale earnings of each case relative to the relevant poverty line which in this figure was inferred from the grant amount. Our measure of AU size is based upon the assumption that size has not changed since the last time the woman was on welfare and had an unreduced grant amount.[11] The close correspondence at low earnings levels between the median monthly grant (which includes cases not on welfare) and the statutory unreduced grant amount indicates that grants are rarely reduced.

Figure 2 strongly suggests that the JF eligibility notch was enforced, with the fraction of women on assistance throughout the quarter falling quickly as earnings exceed the eligibility notch at three times the monthly poverty line. This drop is equally pronounced in several different measures of the duration of welfare participation which suggests lags between earnings and eligibility decisions are not particularly important.

[10] We drop one AFDC case from our analysis with unrealistically high quarterly earnings that sometimes led to erratic results.

[11] We have also tried throwing out cases that experience a change in AU size and found similar results.

We interpret the fact that participation does not drop to zero immediately above the notch as evidence that women substantially under-report their earnings. This is consistent with the analysis of Hotz, Mullin, and Scholz (2003) who analyzed data from a welfare reform experiment in California. Comparing administrative earnings records from the California unemployment insurance system to earnings reported to welfare over a similar time period, they find a highly nonlinear reporting relationship with cases having quarterly UI earnings above $2500 being disproportionately more likely to under-report and under-reporting greater amounts on average. This threshold is somewhat below the federal poverty line for a typical AU and close to the eligibility threshold of the California program. In section 6, we more carefully consider how to distinguish under-reporting responses from true earnings responses.

4 Impacts on Earnings and Participation

We turn now to examining the qualitative predictions of the standard neoclassical model in section 2. Figure 3 provides reweighted CDFs of earnings in the AFDC and JF groups over the seven quarters following random assignment. We rescale earnings relative to three times the monthly FPLs faced by these women based upon our survey measure of AU size ($3FPL$ is the maximum amount that can be earned in a quarter while maintaining welfare eligibility under JF throughout the quarter). To avoid understating the location of the poverty line these women face, we inflate the implied AU size by one.[12]

Several of the qualitative predictions of the theory can be assessed from this figure alone. A reweighted Kolmogorov-Smirnov test strongly rejects the null hypothesis that the two CDFs are identical.[13] Clearly more women had quarters with positive earnings in the JF sample than under AFDC, indicating that JF successfully incentivized many women to work. The earnings CDF rises more quickly in the JF sample than under AFDC, signaling excess mass at low earnings levels. Also, the CDFs cross below the notch suggesting a small opt-in response with the fraction earning below $3FPL$ slightly greater for the JF sample than among the AFDC controls. There is no evidence of a spike in the distribution of earnings at $3FPL$, an issue we will explore more carefully in the next section using a different measure of AU size.

[12] That is, we use the following mapping from kidcount to AU size: 0–>3, 1–>3, 2–>4, 3–>5. This mapping is conservative in ensuring that earnings levels below the FPL are indeed below it. We have found that our qualitative results are relatively robust to alternate codings including inflating the AU size by two and not inflating it at all.

[13] See the appendix for details on the computation of this test.

These distributional effects conceal substantial heterogeneity across subgroups. Figures 4a-4c provide corresponding CDFs in three subsamples considered by BGH (2010) defined by their earnings in the seventh quarter prior to random assignment.[14] These groups are of interest because pre-random assignment earnings are a strong predictor of post-random assignment earnings, hence they proxy for the relevant range of the budget set an agent would face under AFDC. Cases with high pre-random assignment earnings are most likely to exhibit an opt-in effect while cases with zero earnings are likely to be pushed into the labor force by JF. The figures confirm that the expected pattern of heterogeneity is in fact present, with the high earnings group experiencing no impact on the fraction of cases working but a large (though only marginally significant) impact on the fraction of cases with earnings less than or equal to three times the monthly poverty line. The zero earnings group, by contrast, exhibits a large impact on the fraction of cases working, but essentially no impact on the fraction of cases with earnings less than or equal to three times the monthly poverty line.

Table 3 quantifies the distributional impacts of JF on the CDF of earnings and provides standard errors generally confirming the visual impression of the prior figures. JF yielded a large decrease in the fraction of quarters with earnings below the monthly poverty line, suggesting many workers were incentivized to work greater amounts than before by the reduction in implicit tax rates. We also see a tendency for the fraction earning less than three times the poverty line to increase, especially in the higher earning groups, suggesting the presence of an opt-in effect.

Consistent with the model's other predictions, welfare participation grew sharply under JF in each group and the fraction of quarters spent on welfare with no earnings fell. We also see that average earnings while on welfare grew, which is likely driven by a combination of non-random selection into welfare and substitution effects among working woman who would participate in welfare under either set of policy rules.

5 Bunching Evidence

Our evidence so far strongly suggests the presence of an opt-in response to the JF intervention, with the magnitude of the response being greatest in subsamples of the population where one would most expect it. The basic neoclassical model in section 2 predicts that such responses should be accomplished in part via bunching at the eligibility notch. Here

[14] The seventh quarter prior to random assignment is a useful stratifying variable because welfare applicants and recipients generally experience a severe dip in earnings in the quarters immediately prior to random assignment.

we examine this prediction more carefully by looking at bunching in the cross-sectional distribution of earnings in the JF sample where we have a means of accurately measuring AU size.

Figure 5 provides a histogram of earned income rescaled relative to the federal poverty line. Not only do we not see a spike in the mass of cases located at the notch, the earnings density actually appears to be declining through this point. Moreover, this decline is relatively smooth through the notch which, to its right, should be a dominated earnings region for most households. However, compared to individuals not on welfare at all in the quarter there is arguably an excess "mound" in the density of earnings below the notch.

One explanation for these findings is that agents cannot fully control their earnings (Heckman, 1983; Saez, 2010; Chetty et al, 2011). For example, women may engage in directed search for jobs with earnings near the notch level. If a woman receives an offer paying a few hundred dollars less than the notch, she may take it, particularly if she is credit constrained, as the cost of searching for an ideal offer is too great. We also suspect that under-reporting behavior is important for these patterns, as women may report earnings at the notch while earning higher amounts, placing them in what appears to be the dominated region of earnings. The next section develops an augmented model that nests these explanations.

6 An Augmented Model

Thus far we have found that most of the qualitative predictions of the frictionless static model regarding earnings and program participation appear to hold in the data but that predictions regarding a mass point at the eligibility notch fail. We also found that many seemingly ineligible households receive welfare benefits. We consider now a generalized version of the standard labor supply model of section 2 that can accommodate these observations by allowing constraints on worker choices and under-reporting behavior. Working with this model, we derive in the next section an approach to identification of the magnitude of the intensive margin opt-in effect, the extensive margin decision to work, and various sorts of under-reporting responses.

We begin by assuming that, in addition to the welfare participation decision (D) and the choice of earned income (E), workers may also choose a level of earnings (E^r) to report to the welfare agency. We assume that women can under-report but not over-report earnings so that $E^r \leq E$.[15] Grant amounts are determined based upon reported earnings, so that the

[15] Allowing over-reporting behavior would essentially nullify the JF work requirements. In practice, concocting a fictitious job would have been difficult as employment had to be verified by case workers.

transfer function becomes:

$$\begin{aligned} G_i^a(E^r) &= \max\left\{\overline{G}_i - 1[E^r > \delta_i](E^r - \delta_i)\tau_i, 0\right\} \\ G_i^j(E^r) &= 1[E^r < FPL_i]\overline{G}_i. \end{aligned} \tag{3}$$

The costs of under-reporting earnings are given by $\kappa_i \geq 0$ which may vary arbitrarily across women. One can think of κ_i as the psychic and pecuniary costs of concealing a job from the welfare agency.[16] Like stigma and hassle, under-reporting costs are assumed to enter as dollar equivalents in consumption. Thus, we augment our earlier specification of the consumption equivalent to be:

$$C = E + \left(G_i^t(E^r) - \phi_i - \eta_i^t 1[E^r = 0] - \kappa_i 1[E^r < E]\right) D - \mu_i 1[E > 0]. \tag{4}$$

Note that women face hassle when they report zero earnings regardless of whether or not their actual earnings are zero. Hence, a woman concealing an income source will choose to report positive earnings equal to or below the fixed disregard level δ_i.

We make one additional assumption on psychic costs relative to section 2 which will greatly simplify our analysis in the next section: we assume that monthly welfare stigma ϕ_i is bounded from below for women eligible for the proportional disregard. Specifically, we assume $\phi_i \geq G_i^a(FPL_i)$. This assumption guarantees that women will not choose to report earnings above the federal poverty line while on AFDC (e.g. point C in Figure 1).[17] Because under AFDC the proportional disregard phases out just above the poverty line, this bound is not very stringent – at most it requires a monthly stigma of \$75.[18] Moreover, most of the women in our sample are long term participants for whom the assumption is irrelevant because they are ineligible for the proportional disregard.

To model constraints, we suppose that each woman draws a pair of earnings offers (O_i^1, O_i^2) from the bivariate distribution $F_i(.)$. She can choose between these offers or reject them both, in which case she earns nothing. The offers drawn are invariant to the policy regime t to

[16] See Saez (2010) for another analysis of fixed reporting costs.

[17] This restriction is useful for two reasons. First, it allows us to identify women on assistance with earnings above the poverty line as under-reporters regardless of the policy regime (AFDC or JF). Second, it allows us to rule out the possibility that women are induced by JF to stop working. This might happen if the only earnings offers received were above the poverty line, in which case working on welfare would not be an option under JF.

[18] As a check on the plausibility of this assumption we examined the fraction of women in the AFDC sample who participated in welfare while receiving a rounded benefit of \$50. We found this fraction to be approximately 1% which corroborates the notion that most women in our sample do not find it worthwhile to participate in welfare in exchange for very low benefit levels.

which the woman is assigned. Thus, the woman's objective is to:

$$\max_{E\in\left\{0,O_i^1,O_i^2\right\},\ D\in\{0,1\},\ E^r\in[0,E]} U_i\left(E,C\right) \text{ subject to } (3) \text{ and } (4)$$

Note that the presence of two offers provides the possibility of an intensive margin earnings response to the policy change. This response may or may not be constrained as the offers (O_i^1, O_i^2) could coincide with the woman's unconstrained choices of E_i under the two policy regimes. The presence of constraints provides a simple explanation for the absence of a spike in the earnings distribution at the JF eligibility notch. Likewise, our allowance for under-reporting behavior provides an explanation for welfare participation among households with administrative earnings above the eligibility notch.

The introduction of under-reporting behavior introduces new margins of adjustment not present in the model of section 2. Figure 6 illustrates the decision problem in earnings and consumption equivalent space for a woman with substantial fixed costs of work and low costs of under-reporting. The effective budget sets are discontinuous at zero earnings due to the fixed costs of work μ and hassle costs. In the Figure, the hassle costs η^j of not working under JF are substantially larger than those under AFDC represented by η^a, but both are smaller than μ. In comparison with the fixed costs of work and hassle, the costs of under-reporting (κ) are depicted as being relatively small. The under-reporting line is equivalent under AFDC and JF because in either case the woman can report earnings arbitrarily close to zero, obviating any implicit taxes on her true earnings.

A women with this configuration of psychic costs who under AFDC would receive benefits without working (point A) may under JF choose to earn above the poverty line while under-reporting earnings (e.g. point B) in order to maintain eligibility. This occurs because the JF work requirements effectively remove point A from the budget set due to increased hassle. With preferences of the sort depicted in the Figure, options like point B will be an attractive response if the woman draws an earnings offer in that range. Conversely, a woman who works while on AFDC but under-reports earnings (e.g. point C) may choose under JF to work and truthfully report her earnings (e.g. point D) as the earnings disregard reduces the return to under-reporting. Thus, the JF reform has mixed effects on reporting behavior which may lead to an increase or a decrease in the total rate of under-reporting. In the next section we seek to formally identify the magnitude of these and other responses.

7 Identification and Estimation of Response Margins

Our augmented model is sufficiently general that it places no restrictions on the cross-sectional distribution of earnings and program participation choices under a given policy regime. There is a mix of preferences and earnings opportunities that can support any earnings and program participation choice. Hence, the right mix of preferences and offers across women could support any cross-sectional distribution of choices.

However, as we have already discussed, the model does yield predictions about how different groups of women may respond to policy variation. In particular, it rules out transitions between certain states of the world defined by broad earnings and program participation categories. For example, a woman who would have worked on welfare under AFDC with earnings below the eligibility notch will not choose to stop working under JF. This conclusion follows from a revealed preference argument – she could have not worked under AFDC and received the same grant amount (along with less hassle). Since she did not, she must prefer working to this choice.

In this section we enumerate a complete set of restrictions on transitions between coarsened earnings categories and welfare participation states. Our focus on coarsened earnings categories simplifies our analysis and highlights the fact that the experiment only differentially manipulates the relative attractiveness of labor supply choices across broad earnings ranges.[19] We use these restrictions to test the model and form bounds on the fraction of women in each state who make each of the allowable transitions to other states.

Response margins

We begin by introducing some notation. Our coarsened earnings variable $\widetilde{E}_i$ is defined by the relation:

$$\widetilde{E}_i = \begin{cases} 0 & \text{if } E_i = 0 \\ 1 & \text{if } E_i \leq FPL_i \\ 2 & \text{if } E_i > FPL_i \end{cases} .$$

That is, it indicates whether a woman works, and if so, whether she earns enough to be ineligible for benefits under JF. This choice of earnings categories is crucial as the model

[19] A coarsening of the choice set is common practice in the structural labor supply literature (e.g. Hoynes, 1996; Keane and Moffitt, 1998). Typically, the coarsening adopted in these works rests on data-driven categories such as part-time and full-time work. Our approach is fundamentally different. We do not assume that agents lack the ability to choose earnings levels within earnings categories – constraints in our framework are summarized by the earnings offer distribution $F_i(.)$ which may be continuous. Moreover, we coarsen the outcomes based upon features of the policy rules, breaking earnings into regions defined by their relation to the JF eligibility notch. This choice allows us to make strong nonparametric predictions about behavior.

rules out many transitions between (but not within) these categories in response to the JF reform. Moreover, our assumptions so far imply that, under either policy regime, a woman with $\widetilde{E}_i = 2$ who participates in welfare must be under-reporting her earnings to the welfare agency.

Intersecting these earnings categories with the decision to participate in welfare and the under-reporting decision there are seven possible earnings / participation / reporting states allowed by the model:

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

The number of each state refers to the woman's earnings category, while the letter n denotes welfare non-participation, r denotes participating in welfare while truthfully reporting earnings, and u denotes participating in welfare while under-reporting earnings.

Let S_i^a denote woman i's potential state under AFDC and S_i^j her potential state under JF. Define the conditional probability of occupying state $s^j \in \mathcal{S}$ under JF given the choice of state $s^a \in \mathcal{S}$ under AFDC as:

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

These transition probabilities will serve as our parameters of interest as they summarize the frequency of adjustment along the various margins through which agents can respond to the JF reform.

Our model implies the 7×7 matrix of state transition probabilities may be written:

State under AFDC	Earnings / Reporting State under Jobs First						
	$0n$	$1n$	$2n$	$0r$	$1r$	$1u$	$2u$
$0n$	$1-\pi_{0n,1r}$	0	0	0	$\pi_{0n,1r}$	0	0
$1n$	0	$1-\pi_{1n,1r}$	0	0	$\pi_{1n,1r}$	0	0
$2n$	0	0	$1-\pi_{2n,1r}$	0	$\pi_{2n,1r}$	0	0
$0r$	$\pi_{0r,0n}$	0	$\pi_{0r,2n}$	$1-\pi_{0r,0n}-\pi_{0r,2n}$ $-\pi_{0r,1r}-\pi_{0r,2u}$	$\pi_{0r,1r}$	0	$\pi_{0r,2u}$
$1r$	0	0	0	0	1	0	0
$1u$	0	0	0	0	1	0	0
$2u$	0	0	0	0	$\pi_{2u,1r}$	0	$1-\pi_{2u,1r}$

A proof that the matrix takes this form is given in the Appendix. Note that each of the allowable transitions corresponds to a response discussed in Sections 2 or 6. Key to identification of the transition probabilities are the many transitions that the model says cannot occur in response to the JF intervention. Each of these transitions can be ruled out by revealed preference arguments (details in the Appendix).

The model allows eight response margins – that is, eight ways in which a woman might respond to the JF experiment by changing her behavior. Reading the above matrix from the top left element to the bottom right element, these margins are:

- A fraction $\pi_{0n,1r}$ of women who would not participate in welfare or work under AFDC will take up welfare and work under JF.
- A fraction $\pi_{1n,1r}$ of women who would work but earn less than the poverty line and not participate in welfare under AFDC will work while on welfare under JF.
- A fraction $\pi_{2n,1r}$ of women who would earn more than the poverty line and be off assistance under AFDC will "opt-in" to welfare and reduce their earnings below the poverty line under JF.
- A fraction $\pi_{0r,0n}$ of women who participate in welfare without working under AFDC will leave welfare and continue not to work.
- A fraction $\pi_{0r,2n}$ of women who participate in welfare without working under AFDC will leave welfare and earn above the poverty line.
- A fraction $\pi_{0r,1r}$ of women who participate in welfare without working under AFDC will take up work and remain on welfare under JF.
- A fraction $\pi_{0r,2u}$ of women who participate in welfare without working under AFDC will earn above the poverty line but remain on welfare by under-reporting earnings.
- A fraction $\pi_{2u,1r}$ of women who earn above the poverty line but under-report earnings in order to qualify for benefits under AFDC will reduce their earnings below the poverty line under JF and truthfully report earnings.

Note that some of these transitions ($\pi_{0r,2n}, \pi_{0r,2u}$) involve moving from earnings category 0 to category 2, while others ($\pi_{2n,1r}, \pi_{2u,1r}$) involve moving from earnings category 2 to category 1. Thus, the model allows for rank reversals in earnings and therefore violates the standard rank-invariance condition.

A variety of other adjustments may occur within each of our three earnings ranges. Thus the restriction that $\pi_{1p,1p} = 1$ should not be taken to imply that no response is present among women who work on welfare, as many such women may increase their earnings. Without further restrictions, however, we cannot infer the magnitude of any such adjustments.

Identification

Identification of the transition probabilities is hindered by the fact that the states $\{1u, 1r\}$ in $\mathcal{S}$ are not empirically distinguishable. The function $g(s) : \mathcal{S} \rightarrow \widetilde{\mathcal{S}}$ defined as

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

maps the set of latent states $\mathcal{S}$ to the set of observed states $\widetilde{\mathcal{S}} \equiv \{0n, 1n, 2n, 0p, 1p, 2p\}$. As before, the number of each state refers to the woman's earnings category and the letter n refers to welfare non-participation. The letter p denotes welfare participation, which is directly observable. Note that state $2p$ can only be occupied via under-reporting.

Let $\widetilde{S}_i$ denote the value of woman i's *observed* state, $\widetilde{S}_i^a = g(S_i^a)$ her potential observed state under AFDC, and $\widetilde{S}_i^j = g\left(S_i^j\right)$ her potential observed state under JF. The experimental status variable $T_i \in \{a, j\}$ denotes the policy regime into which woman i was randomized. Random assignment of women to the two policy regimes guarantees that the following condition holds:

$$T_i \perp \left(S_i^a, S_i^j\right) \tag{5}$$

where the $\perp$ symbol denotes independence.

Define

$$p_{\widetilde{s}}^t \equiv P\left(\widetilde{S}_i^t = \widetilde{s}\right) = \sum_{s:\ g(s)=\widetilde{s}} P\left(S_i^t = s\right)$$

as the probability of occupying observable state $\widetilde{s} \in \widetilde{\mathcal{S}}$ under treatment regime t. The marginal state probabilities $p_{\widetilde{s}}^t$ are identified by the relation

$$p_{\widetilde{s}}^t = P\left(\widetilde{S}_i = \widetilde{s} | T_i = t\right),$$

which follows from (5). Thus, we have the well-known result that experimental variation identifies the marginal distributions of potential outcomes.

Define the vectors of observable state probabilities:

$$\begin{aligned}\mathbf{p}^j &\equiv \left[p^j_{0n}, p^j_{1n}, p^j_{2n}, p^j_{0p}, p^j_{1p}, p^j_{2p}\right]' \\ \mathbf{p}^a &\equiv \left[p^a_{0n}, p^a_{1n}, p^a_{2n}, p^a_{0p}, p^a_{1p}, p^a_{2p}\right]'.\end{aligned}$$

By the law of total probability,

$$\mathbf{p}^j = \mathbf{\Pi}\mathbf{p}^a \tag{6}$$

where the 6×6 transition matrix $\mathbf{\Pi}$ is composed of terms of the form $P\left(\widetilde{S}^j_i = \widetilde{s}^j | \widetilde{S}^a_i = \widetilde{s}^a\right)$. Specifically, $\mathbf{\Pi}$ may be written:[20]

State under AFDC	Earnings / Participation State under Jobs First					
	$0n$	$1n$	$2n$	$0p$	$1p$	$2p$
$0n$	$1-\pi_{0n,1r}$	0	0	0	$\pi_{0n,1r}$	0
$1n$	0	$1-\pi_{1n,1r}$	0	0	$\pi_{1n,1r}$	0
$2n$	0	0	$1-\pi_{2n,1r}$	0	$\pi_{2n,1r}$	0
$0p$	$\pi_{0r,0n}$	0	$\pi_{0r,2n}$	$1-\pi_{0r,0n}-\pi_{0r,2n}$ $-\pi_{0r,1r}-\pi_{0r,2u}$	$\pi_{0r,1r}$	$\pi_{0r,2u}$
$1p$	0	0	0	0	1	0
$2p$	0	0	0	0	$\pi_{2u,1r}$	$1-\pi_{2u,1r}$

.

The system in (6) consists of six equations, one of which is redundant given that state probabilities sum to one in each policy regime. The five non-redundant equations can be given an intuitive representation as:

$$\begin{aligned} p^j_{0n} - p^a_{0n} &= -p^a_{0n}\pi_{0n,1r} + p^a_{0p}\pi_{0r,0n} \\ p^j_{1n} - p^a_{1n} &= -p^a_{1n}\pi_{1n,1r} \\ p^j_{2n} - p^a_{2n} &= -p^a_{2n}\pi_{2n,1r} + p^a_{0p}\pi_{0r,2n} \\ p^j_{0p} - p^a_{0p} &= -p^a_{0p}\left(\pi_{0r,1r} + \pi_{0r,2u} + \pi_{0r,2n} + \pi_{0r,0n}\right) \\ p^j_{2p} - p^a_{2p} &= p^a_{0p}\pi_{0r,2u} - p^a_{2p}\pi_{2u,1r} \end{aligned} \tag{7}$$

The left hand side of (7) catalogues the experimental impacts of the JF reform on the observable state probabilities. The right hand side catalogues the flows into and out of each

[20] Note that these transition probabilities do not depend on the proportion of women $P\left(S^a_i = \{1u\}\right)$ who would earn below the poverty line yet under-report their earnings under AFDC policy rules. This occurs because women who would work and earn less than the poverty line under AFDC rules are predicted to do the same under Jobs First, regardless of whether under AFDC they would have under-reported their earnings.

state as allowed by the labor supply model. The identifying power of the theory derives from the fact that only a handful of transition probabilities appear in each equation. Despite these restrictions, the system in (7) is clearly under-determined, with eight unknown transition probabilities and only five equations. Still, it is immediate to see that the second equation of (7) uniquely identifies the transition probability $\pi_{1n,1r}$. The remaining 4 equations constrain (without uniquely determining) the remaining seven transition probabilities.

Some of the restrictions embedded in (7) are testable. For instance, the model implies that the experiment cannot induce entry into states $1n$ or $0p$. The complete set of testable restrictions is:

$$\begin{aligned} p^a_{0p} - p^j_{0p} &\geq 0 \\ p^j_{1p} - p^a_{1p} &\geq p^a_{1n} - p^j_{1n} \geq 0 \end{aligned} \tag{8}$$

Violation of any of these conditions would imply that our framework failed to allow a response actually present in the data.

Subject to the restrictions in (8) holding, we can use the system in (7) to bound the seven remaining transition probabilities. The upper and lower bounds on each of the transition probabilities can be represented as the solution to a pair of linear programming problems of the form:

$$\begin{aligned} &\max_{\boldsymbol{\pi}} \boldsymbol{\lambda}'\boldsymbol{\pi} \\ \text{s.t. } \boldsymbol{\pi} &\in [0,1]^7 \text{ and } (7) \end{aligned} \tag{9}$$

where $\boldsymbol{\pi} \equiv [\pi_{0n,1r}, \pi_{0r,0n}, \pi_{2n,1r}, \pi_{0r,2n}, \pi_{0r,1r}, \pi_{0r,2u}, \pi_{2u,1r}]'$. For example, solving the above problem for $\boldsymbol{\lambda} = [0,0,0,0,0,0,1]'$ yields the upper bound on $\pi_{2u,1r}$, while choosing $\boldsymbol{\lambda} = [0,0,0,0,0,0,-1]'$ yields the lower bound.

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

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

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

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

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

The parameter $\pi_{0r,n}$ gives the fraction of women who would claim benefits without working under AFDC that are induced to get off welfare under JF. Upper and lower bounds for this transition probability can be had by solving (9) with $\boldsymbol{\lambda} = [0,1,0,1,0,0,0]$ and $[0,-1,0,-1,0,0,0]$ respectively. We also examine the fraction $\pi_{p,n}$ of all women who would participate in welfare under AFDC that are induced to leave welfare under JF, the fraction $\pi_{n,p}$ of women who are induced to take up welfare under JF, and the fraction $\pi_{0,1+}$ who are induced by JF to work. Because no women 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.

To conduct inference, it is convenient to derive analytic expressions for the bounds as a function of the regime-specific marginal distributions entering the constraints in (9). We accomplished this by solving the relevant linear programming problems by hand (a straightforward though cumbersome process). The resulting expressions are listed in the Appendix. An example is given by the bounds on the opt-in probability $\pi_{2n,1r}$ which obey:

$$\max\left\{0, \frac{p^a_{2n}-p^j_{2n}}{p^a_{2n}}\right\} \leq \pi_{2n,1r} \leq \min\left\{\begin{array}{c} 1, \\ \frac{p^a_{2n}-p^j_{2n}+p^a_{0p}-p^j_{0p}}{p^a_{2n}}, \\ \frac{p^a_{2n}-p^j_{2n}+p^a_{0p}-p^j_{0p}+p^a_{0n}-p^j_{0n}}{p^a_{2n}}, \\ \frac{p^a_{2n}-p^j_{2n}+p^a_{0p}-p^j_{0p}+p^a_{2p}-p^j_{2p}}{p^a_{2n}}, \\ \frac{p^a_{2n}-p^j_{2n}+p^a_{0p}-p^j_{0p}+p^a_{0n}-p^j_{0n}+p^a_{2p}-p^j_{2p}}{p^a_{2n}} \end{array}\right\}.$$

Note that there are two possible solutions for the lower bound, one of which is zero. This turns out to be a generic feature of the lower bounds for each of the seven transition probabilities. Which solution will be relevant is unknown a priori. The upper bound on $\pi_{2n,1r}$ admits five possible solutions. Other transition probabilities can have fewer or more solutions.

Estimation and Inference

Consistent estimators of the upper and lower bounds can be had by using sample analogs of the marginal probabilities and computing the relevant $\min\{.\}$ and $\max\{.\}$ expressions. 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 (9) bind – i.e. in which of the bound solutions is relevant. Naive bootstrap inference on the empirical $\min\{.\}$ and $\max\{.\}$ of the sample analogues of the bound solutions will fail to provide coverage of the parameters in question with fixed probability (Andrews and Han, 2009).

We report confidence intervals for the transition probabilities based upon two inference procedures. The first simply ignores the uncertainty in which constraints bind – that is, it assumes the bound solution that appears relevant given the sample analogues is the only possible solution. In such a case, results from Imbens and Manski (2002) imply a 95% confidence interval for the parameter in question can be constructed by extending the upper and lower bounds by $1.65\widehat{\sigma}$ where $\widehat{\sigma}$ is a standard bootstrap estimate of the standard error of the sample moment used to define the relevant bound.[21] The second approach is a conservative bootstrap procedure described in the Appendix which covers the parameter with asymptotic probability greater than or equal to 95% regardless of which constraints bind. The lower limit of this confidence interval coincides with that of the naive procedure because sampling uncertainty only affects one of the bound solutions in the $\max\{.\}$ operator. However, the upper limit of the confidence interval from our conservative procedure generally exceeds that from the naive procedure, often by a substantial amount.

8 Results

Table 4 reports estimated probabilities of occupying the six observable earnings and welfare participation states under each policy regime. Notably, the sign restrictions in (8) are satisfied by the point estimates. We also see a small but statistically significant increase in the fraction of quarters on welfare with earnings above the quarterly poverty line indicating that some women were induced by the treatment to under-report their earnings.

Table 5 provides estimates of the transition probabilities that rationalize the impacts in Table 4. The point identified transition probability $\pi_{1n,1r}$ is computed by plugging its sample analogue. We find a strong effect of JF on entry into the program by the working poor. Our bootstrap confidence intervals suggest between 31% and 46% of the women who would have worked off welfare under the AFDC system at earnings levels below the poverty line were induced to participate in JF at eligible earning levels.

[21] For example, if the relevant lower bound for $\pi_{2n,1r}$ is $\frac{p^a_{2n}-p^j_{2n}}{p^a_{2n}}$ and the relevant upper bound is $\frac{p^a_{2n}-p^j_{2n}+p^a_{0p}-p^j_{0p}}{p^a_{2n}}$, then the 95% bootstrap confidence interval for $\pi_{2n,1r}$ is: $\left[\max\left\{0,\frac{\widehat{p}^a_{2n}-\widehat{p}^j_{2n}}{\widehat{p}^a_{2n}}-1.65\widehat{\sigma}_l\right\},\min\left\{1,\frac{\widehat{p}^a_{2n}-\widehat{p}^j_{2n}+\widehat{p}^a_{0p}-\widehat{p}^j_{0p}}{\widehat{p}^a_{2n}}+1.65\widehat{\sigma}_u\right\}\right]$ with $\widehat{\sigma}_l$ the bootstrap standard error of $\frac{\widehat{p}^a_{2n}-\widehat{p}^j_{2n}}{\widehat{p}^a_{2n}}$ and $\widehat{\sigma}_u$ the bootstrap standard error of $\frac{\widehat{p}^a_{2n}-\widehat{p}^j_{2n}}{\widehat{p}^a_{2n}}+\frac{\widehat{p}^a_{0p}-\widehat{p}^j_{0p}}{\widehat{p}^a_{2n}}$ and where hats on probabilities denote reweighted sample analogues.

We also find a substantial opt-in response among women who would have worked off welfare at earning levels above the poverty line. Our estimated bounds imply that $\pi_{2n,1r} \geq .28$. That is, at least 28% of those women with ineligible earnings under AFDC decided to work at eligible levels under JF and participate in welfare. Accounting for sampling uncertainty in the bounds extends this lower limit to 19%, which is still quite substantial. The upper bounds for this parameter were not informative leading us to conclude that the opt-in probability is somewhere in the interval $[.19, 1]$ with 95% probability.

We also find suggestive evidence of a second opt-in effect from non-participation. Our sample bounds imply $\pi_{0n,1r} \in [.06, .62]$. However, uncertainty in the bounds prevents us from rejecting the null that this response probability is actually zero. We also find a small but significant under-reporting response attributable to the hassle effects of JF. A conservative 95% confidence interval for $\pi_{0r,2u}$ is $[.02, .13]$. Thus, JF induced at least one subpopulation to under-report earnings, and in the process violate the standard rank invariance condition implicit in BGH's analysis.

The remaining transition probabilities $(\pi_{0r,0n}, \pi_{0r,2n}, \pi_{0r,1r}, \pi_{2u,1r})$ each have zero lower bounds. However, we can reject the null that they are jointly zero. From (7) such a joint restriction implies:

$$p_{0p}^{j} - p_{0p}^{a} = -\left(p_{2p}^{j} - p_{2p}^{a}\right)$$

which is easily rejected by 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. However, we cannot be sure that the abundance of women working at low earning levels under JF are in fact coming from state $0r$ rather than state $2u$.

We also computed bounds on the three composite adjustment margins described in the previous section. First is the probability $\pi_{0,1+}$ that a woman transitions along the extensive margin from nonwork to work. A conservative 95% confidence interval for this probability is $[0.13, 0.21]$. Thus JF induced a substantial fraction of women to work who would not have done so under AFDC.

We also find a relatively tight confidence interval on the fraction $\pi_{n,p}$ of women induced to take up welfare by JF. 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% (and at most 51%) 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 least zero and at most 17%.

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 transition probability $\pi_{0r,n}$ is zero. We are however able to conclude that at most 24% 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.

9 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 would traditionally be considered intensive and some of which are clearly extensive. 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.

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 be more salient than, say, the budget kink induced by the EITC phase-out region. However, BGH (2008) show that experimental responses to a Canadian reform inducing such a gradual benefit phaseout 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 generalized to quasi-experimental settings. Given an appropriate coarsening of the budget set, one can form estimates of counterfactual choice probabilities using one's research design of choice (e.g., the difference in differences design adopted by Meyer and Rosenbaum, 2001), subject to the usual caveat that different designs may identify counterfactuals for different treated subpopulations.[22] With the two sets of marginal choice probabilities, bounds on transition probabilities can then be had by a direct application of the methods developed in this paper.

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

As with most methods designed for the study of treatment effects, we cannot, without additional assumptions, predict the responses likely to arise from new interventions outside the range of observed policy variation. In cases where data are available on many different sorts of policy interventions, one can fit a curve summarizing how the transition bounds vary with policy parameters and attempt a statistical extrapolation. Otherwise, restrictions on model primitives will be necessary for prediction. A natural approach would be to parameterize utility and the process governing the labor supply constraints (e.g. as in Chetty et al., 2011), in which case bounds can be developed on structural parameters rather than transition probabilities. We leave the development of such methods to future work.

References

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

2. Andrews, Donald and S. Han. 2009. "Invalidity of the Bootstrap and m Out of n Bootstrap for Interval Endpoints Defined by Moment Inequalities." *Econometrics Journal* 12(s1): S172-S199.

3. Andrews, Donald and Panle Jia Barwick. 2012. "Inference for Parameters Defined by Moment Inequalities: A Recommended Moment Selection Procedure." *Econometrica* 80(6): 2805–2826.

4. Ashenfelter, Orley. 1983. "Determining Participation in Income-Tested Social Programs." *Journal of the American Statistical Association* 78(383):517-525.

5. Bitler, Marianne, Jonah Gelbach, and Hilary Hoynes. 2005. "Distributional Impacts of the Self-Sufficiency Project." *NBER Working Paper #11626.*

6. Bitler, Marianne, Jonah Gelbach, and Hilary Hoynes. 2008. "Distributional Impacts of the Self-Sufficiency Project." *Journal of Public Economics 92:748-765.*

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

8. Bitler, Marianne, Jonah Gelbach, and Hilary Hoynes. 2010. "Can Variation in Subgroups' Average Treatment Effects Explain Treatment Effect Heterogeneity? Evidence from a Social Experiment." working paper.

9. Blundell, Richard and Thomas Macurdy. 1999. "Labor Supply: A Review of Alternative Approaches." In *Handbook of Labor Economics*, Vol. 3A, ed. Orley Ashenfelter and David Card, 1559–1695. Amsterdam: Elsevier.

10. 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.

11. Blundell, Richard, Antoine Bozio, and Guy Laroque. 2011b. "Labour Supply and the Extensive Margin." *American Economic Review, Papers and Proceedings* 101(3): 482-486.

12. Burtless, Gary and Jerry Hausman. 1978. "The Effect of Taxation on Labor Supply: Evaluating the Gary Negative Income Tax Experiment." *The Journal of Political Economy* 86(6):1103-1130.

13. Card, David and Dean Hyslop. 2005. "Estimating the Effects of a Time-Limited Earnings Subsidy for Welfare-Leavers." *Econometrica*, 73(6): 1723-1770.

14. Chetty, Raj, John 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.

15. Chetty, Raj, John Friedman, and Emmanuel Saez. 2012. "Using Differences in Knowledge Across Neighborhoods to Uncover the Impacts of the EITC on Earnings." NBER Working Paper #18232.

16. Chernozhukov, Victor, Sokbae Lee, and Adam Rosen. 2009. "Intersection bounds: estimation and inference." working paper.

17. Deaton, Angus. 2009. "Instruments of Development: Randomization in the tropics, and the search for the elusive keys to economic development." NBER Working Paper #14690.

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

19. Eissa, Nada, and Hilary Hoynes. 2006. "The Hours of Work Response of Married Couples: Taxes and the Earned Income Tax Credit." In *Tax Policy and Labor Market Performance*, ed. Jonas Agell and Peter Birch Sorensen, 187–227. Cambridge: MIT Press.

20. Ferrall, Christopher. 2012. "Explaining and Forecasting Results of The Self-Sufficiency Project." *Review of Economic Studies*, forthcoming.

21. Giné, Evarist and Joel Zinn. 1990. "Bootstrapping General Empirical Measures." *Annals of Probability* 18(2): 851-869.

22. Greenberg, David, Robert Moffitt, and John Friedmann. 1981. "Underreporting and Experimental Effects on Work Effort: Evidence from the Gary Income Maintenance Experiment." *Review of Economics and Statistics* 63(4): 581-589.

23. Greenberg, David and Harlan Halsey. 1983. "Systematic Misreporting and Effects of Income Maintenance Experiments on Work Effort: Evidence from the Seattle-Denver Experiment." *Journal of Labor Economics* 1(4): 380-407.

24. Grogger, Jeffrey and Charles Michalopoulos. 2003. "Welfare Dynamics under Time Limits." *Journal of Political Economy* 111(3): 530-554.

25. Hansen, Gary D. 1985. "Indivisible Labor and the Business Cycle." *Journal of Monetary Economics* 16:309–327.

26. Hausman, Jerry. 1985. "The Econometrics of Nonlinear Budget Sets." *Econometrica* 53(6): 1255-1282.

27. Heckman, James and Jeffrey Smith. 1995. "Assessing the Case for Social Experiments." *Journal of Economic Perspectives* 9(2): 85-110.

28. Heckman, James, 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.

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

30. Heckman, James. 1993. "What Has Been Learned About Labor Supply in the Last 20 Years?" *American Economic Review Papers and Proceedings* 83(2): 116-121.

31. Heckman, James. 2001. "Micro Data, Heterogeneity, and the Evaluation of Public Policy: Nobel Lecture." *Journal of Political Economy* 109(4): 673-748.

32. Heim, Bradley and Bruce Meyer. 2004. "Work Costs and Non-Convex Preferences in the Estimation of Labor Supply Models." *Journal of Public Economics* 88(11): 2323-2338.

33. Hotz, Joseph , Charles Mullin, and Karl Scholz. 2003. "The Effects of Welfare Reform on Employment and Income: Evidence from California." working paper.

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

35. Imbens, Guido and Donald Rubin. 1997. "Estimating Outcome Distributions for Compliers in Instrumental Variables Models." *Review of Economic Studies* 64(4): 555–574.

36. Katta, Murty. 1983. *Linear Programming* New York: John Wiley & Sons.

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

38. Kosorok, Michael. 2008. *Introduction to Empirical Processes and Semiparametric Inference* Springer: New York.

39. Laroque, Guy. 2005. "Income Maintenance and Labor Force Participation." *Econometrica* 73(2): 341-376.

40. Ljungqvist, Lars and Thomas Sargent. 2009. "A Labor Supply Elasticity Accord?" *American Economic Review Papers and Proceedings* 101(3) : 487

41. Macurdy, Thomas, David Green, and Harry Paarsch. 1990. "Assessing Empirical Approaches for Analyzing Taxes and Labor Supply." *Journal of Human Resources* 25(3): 415-490.

42. Meyer, Bruce and Dan Rosenbaum. 2001. "Welfare, the Earned Income Tax Credit, and the Labor Supply of Single Mothers." *Quarterly Journal of Economics* 116(3): 1063-1114.

43. Meyer, Bruce. 2002. "Labor Supply at the Intensive and Extensive Margins: the EITC, Welfare, and Hours Worked." *American Economic Review* 92(2): 373-379.

44. Moffitt, Robert. 1990. “The Econometrics of Kinked Budget Constraints.” *Journal of Economic Perspectives* 4(2): 119-139.

45. Moffitt, Robert. 2002. “Welfare Programs and Labor Supply.” In *Handbook of Public Economics*. Vol. 4, ed. Alan J. Auerbach and Martin S. Feldstein, 2393–2430. New York: North-Holland.

46. Rogerson, Richard. 1988. “Indivisible Labor, Lotteries, and Equilibrium.” *Journal of Monetary Economics* 21:3–16.

47. Saez, Emmanuel. 2002. “Optimal Income Transfer Programs: Intensive Versus Extensive Labor Supply Responses.” *Quarterly Journal of Economics* 117(3): 1039-1073.

48. Saez, Emmanuel. 2010. “Do Taxpayers Bunch at Kink Points?” *American Economic Journal: Economic Policy* 2(3): 180-212.

Appendix

Propensity Score Reweighting

To deal with the chance imbalances between the AFDC and JF groups, we estimate a parametric propensity score model. Following BGH (2006) we estimate a logit of the JF assignment dummy on: quarterly earnings in each of the 8 pre-assignment quarters, separate variables representing quarterly AFDC and quarterly food stamps payments in each of the 7 pre-assignment quarters, dummies indicating whether each of these 22 variables is nonzero, and dummies indicating whether the woman was employed at all or on welfare at all in the year preceding random assignment or in the applicant sample. We also include dummies indicating each of the following baseline demographic characteristics: being white, black, or Hispanic; being never married or separated; having a high-school diploma/GED or more than a high-school education; having more than two children; being younger than 25 or age 25–34; and dummies indicating whether baseline information is missing for education, number of children, or marital status.

Denote the predicted values from this model by $\widehat{p}_i$. The propensity score weights used to adjust the moments of interest are given by:

$$\omega_i = \frac{\frac{T_i}{\widehat{p}_i}}{\sum\limits_j \frac{T_j}{\widehat{p}_j}} + \frac{\frac{1-T_i}{1-\widehat{p}_i}}{\sum\limits_j \frac{1-T_j}{1-\widehat{p}_j}}.$$

These are inverse probability weights, re-normalized to sum to one within policy group. When examining subgroups we always recompute a new set of propensity score weights and re-normalize them.

Kolmogorov-Smirnov Tests for Equality of Distribution

We use a bootstrap procedure to compute the p-values for our reweighted Kolmogorov-Smirnov (K-S) tests for equality of distribution functions across treatment groups. Let $F_n^t(e)$ be the reweighted empirical distribution function (EDF) of earnings in treatment group t. That is,

$$F_n^t(e) \equiv \sum_i \omega_i I\left[E_i \leq e, T_i = t\right]$$

Define the corresponding bootstrap EDF as:

$$F_n^{t*}(e) \equiv \sum_i \omega_i^* I\left[E_i^* \leq e, T_i^* = t\right]$$

The K-S test statistic is given by:

$$\widehat{KS} \equiv \sup_e \left| F_n^j(e) - F_n^a(e) \right|$$

To obtain a critical value for this statistic, we compute the bootstrap distribution of the *recentered* K-S statistic:

$$KS^* \equiv \sup_e \left| F_n^{j*}(e) - F_n^{a*}(e) - \left(F_n^j(e) - F_n^a(e)\right) \right|.$$

Recentering is necessary to impose the correct null hypothesis on the bootstrap DGP (Giné and Zinn, 1990). We compute an estimated p-value $\widehat{\alpha}$ for the null hypothesis that the two distributions are equal as:

$$\widehat{\alpha} \equiv \frac{1}{1000} \sum_{b=1}^{1000} I\left[KS^{*(b)} > \widehat{KS}\right]$$

where b indexes the bootstrap replication.

List of Bounds

The analytical expressions for the bounds on the transition probabilities are as follows:

$$
\begin{aligned}
\pi_{2n,1r} &\geq \max\left\{0, \frac{p^a_{2n}-p^j_{2n}}{p^a_{2n}}\right\}, && (10)\\
\pi_{2n,1r} &\leq \min\left\{\begin{array}{c} 1, \frac{p^a_{2n}-p^j_{2n}+p^a_{0p}-p^j_{0p}}{p^a_{2n}}, \frac{p^a_{2n}-p^j_{2n}+p^a_{0p}-p^j_{0p}+p^a_{0n}-p^j_{0n}}{p^a_{2n}}, \\ \frac{p^a_{2n}-p^j_{2n}+p^a_{0p}-p^j_{0p}+p^a_{2p}-p^j_{2p}}{p^a_{2n}}, \frac{p^a_{2n}-p^j_{2n}+p^a_{0p}-p^j_{0p}+p^a_{0n}-p^j_{0n}+p^a_{2p}-p^j_{2p}}{p^a_{2n}} \end{array}\right\}, && (11)
\end{aligned}
$$

$$
\begin{aligned}
\pi_{0n,1r} &\geq \max\left\{0, \frac{p^a_{0n}-p^j_{0n}}{p^a_{0n}}\right\},\\
\pi_{0n,1r} &\leq \min\left\{\begin{array}{c} 1, \frac{p^a_{0n}-p^j_{0n}+p^a_{0p}-p^j_{0p}}{p^a_{0n}}, \frac{p^a_{0n}-p^j_{0n}+p^a_{0p}-p^j_{0p}+p^a_{2n}-p^j_{2n}}{p^a_{0n}}, \\ \frac{p^a_{0n}-p^j_{0n}+p^a_{0p}-p^j_{0p}+p^a_{2p}-p^j_{2p}}{p^a_{0n}}, \frac{p^a_{0n}-p^j_{0n}+p^a_{0p}-p^j_{0p}+p^a_{2p}-p^j_{2p}+p^a_{2n}-p^j_{2n}}{p^a_{0n}} \end{array}\right\},
\end{aligned}
$$

$$
\begin{aligned}
\pi_{2u,1r} &\geq \max\left\{0, \frac{p^a_{2p}-p^j_{2p}}{p^a_{2p}}\right\},\\
\pi_{2u,1r} &\leq \min\left\{\begin{array}{c} 1, \frac{p^a_{2p}-p^j_{2p}}{p^a_{2p}} + \frac{p^a_{0p}-p^j_{0p}}{p^a_{2p}}, \frac{p^a_{2p}-p^j_{2p}+p^a_{0p}-p^j_{0p}+p^a_{2n}-p^j_{2n}}{p^a_{2p}}, \\ \frac{p^a_{2p}-p^j_{2p}}{p^a_{2p}} + \frac{p^a_{0p}-p^j_{0p}+p^a_{0n}-p^j_{0n}}{p^a_{2p}}, \frac{p^a_{2p}-p^j_{2p}+p^a_{0p}-p^j_{0p}+p^a_{2n}-p^j_{2n}+p^a_{0n}-p^j_{0n}}{p^a_{2p}} \end{array}\right\},
\end{aligned}
$$

$$
\begin{aligned}
\pi_{0r,1r} &\geq \max\left\{0, \frac{p^a_{0p}-p^j_{0p}-p^j_{0n}-p^j_{2n}-p^j_{2p}}{p^a_{0p}}\right\},\\
\pi_{0r,1r} &\leq \min\left\{\begin{array}{c} \frac{p^a_{0p}-p^j_{0p}}{p^a_{0p}}, \frac{p^a_{0p}-p^j_{0p}+p^a_{0n}-p^j_{0n}}{p^a_{0p}}, \frac{p^a_{0p}-p^j_{0p}+p^a_{2n}-p^j_{2n}}{p^a_{0p}}, \frac{p^a_{0p}-p^j_{0p}+p^a_{2p}-p^j_{2p}}{p^a_{0p}}, \\ \frac{p^a_{0p}-p^j_{0p}+p^a_{2n}-p^j_{2n}+p^a_{2p}-p^j_{2p}}{p^a_{0p}}, \frac{p^a_{0p}-p^j_{0p}+p^a_{0n}-p^j_{0n}+p^a_{2n}-p^j_{2n}}{p^a_{0p}}, \frac{p^a_{0p}-p^j_{0p}+p^a_{2p}-p^j_{2p}+p^a_{0n}-p^j_{0n}}{p^a_{0p}}, \\ \frac{p^a_{0p}-p^j_{0p}+p^a_{0n}-p^j_{0n}+p^a_{2n}-p^j_{2n}+p^a_{2p}-p^j_{2p}}{p^a_{0p}} \end{array}\right\},
\end{aligned}
$$

$$
\begin{aligned}
\pi_{0r,2n} &\geq \max\left\{0, \frac{-\left(p^a_{2n}-p^j_{2n}\right)}{p^a_{0p}}\right\},\\
\pi_{0r,2n} &\leq \min\left\{\begin{array}{c} \frac{p^j_{2n}}{p^a_{0p}}, \frac{p^a_{0p}-p^j_{0p}}{p^a_{0p}}, \frac{p^a_{0p}-p^j_{0p}+p^a_{2p}-p^j_{2p}}{p^a_{0p}}, \frac{p^a_{0p}-p^j_{0p}+p^a_{0n}-p^j_{0n}}{p^a_{0p}}, \\ \frac{p^a_{0p}-p^j_{0p}+p^a_{2p}-p^j_{2p}+p^a_{0n}-p^j_{0n}}{p^a_{0p}} \end{array}\right\},
\end{aligned}
$$

$$
\begin{aligned}
\pi_{0r,2u} &\geq \max\left\{0, \frac{-\left(p^a_{2p} - p^j_{2p}\right)}{p^a_{0p}}\right\}, \\
\pi_{0r,2u} &\leq \min\left\{\begin{array}{c} \frac{p^j_{2p}}{p^a_{0p}}, \frac{p^a_{0p}-p^j_{0p}}{p^a_{0p}}, \frac{p^a_{0p}-p^j_{0p}+p^a_{2n}-p^j_{2n}}{p^a_{0p}}, \frac{p^a_{0p}-p^j_{0p}+p^a_{0n}-p^j_{0n}}{p^a_{0p}}, \\ \frac{p^a_{0p}-p^j_{0p}+p^a_{2n}-p^j_{2n}+p^a_{0n}-p^j_{0n}}{p^a_{0p}} \end{array}\right\},
\end{aligned}
$$

$$
\begin{aligned}
\pi_{0r,0n} &\geq \max\left\{0, \frac{-\left(p^a_{0n} - p^j_{0n}\right)}{p^a_{0p}}\right\}, \\
\pi_{0r,0n} &\leq \min\left\{\begin{array}{c} \frac{p^j_{0n}}{p^a_{0p}}, \frac{p^a_{0p}-p^j_{0p}}{p^a_{0p}}, \frac{p^a_{0p}-p^j_{0p}+p^a_{2p}-p^j_{2p}}{p^a_{0p}}, \frac{p^a_{0p}-p^j_{0p}+p^a_{2n}-p^j_{2n}}{p^a_{0p}}, \\ \frac{p^a_{0p}-p^j_{0p}+p^a_{2p}-p^j_{2p}+p^a_{2n}-p^j_{2n}}{p^a_{0p}} \end{array}\right\}.
\end{aligned}
$$

Derivation of Bounds

A solution to any linear programming problem has to occur at one of the vertices of the problem's constraint space (see Murty, 1983). To obtain the set of possible solutions to the problem in (9), we enumerated all vertices of the convex polytope defined by the intersection of the hyperplane defined by the equations in (7) with the hypercube defined by the unit constraints on the parameters. In practice, this amounted to setting all possible choices of three of the seven parameters in (7) to 0 or 1 and solving for the remaining four parameters. There were $\binom{7}{3} = 35$ different possible choices of three parameters and $2^3 = 8$ different binary arrangements those parameters could take, yielding 280 possible vertices. However we were able to use the structure of our problem to rule out the existence of solutions at certain vertices – e.g., $\pi_{2n,1r}$ and $\pi_{0r,2n}$ cannot both be set arbitrarily because this would lead to a violation of the second equation in (7). Such restrictions reduced the problem to solving the system at 160 vertices. We then enumerated the set of minima and maxima each parameter could achieve across the 160 relevant solutions. After eliminating dominated solutions, we arrived at the stated bounds.

Inference on Bounds

We begin with a description of the upper limit of our confidence interval. For each transition probability π we have a set of possible upper bound solutions $\{ub_1, ub_2, ..., ub_K\}$. We know that:

$$
\begin{aligned}
\pi &\leq \overline{\pi} \equiv \min\{\underline{ub}, 1\} \\
\underline{ub} &\equiv \min\{ub_1, ub_2, ..., ub_K\}
\end{aligned}
$$

A consistent estimate of the least upper bound $\underline{ub}$ can be had by plugging in consistent sample moments $\widehat{ub}_k \xrightarrow{p} ub_k$ and using $\underline{\widehat{ub}} \equiv min\left\{\widehat{ub}_1, \widehat{ub}_2, ..., \widehat{ub}_K\right\}$ as an estimate of $\underline{ub}$. This estimator is consistent by continuity of probability limits. We can then form a corresponding consistent estimator $\widehat{\overline{\pi}} \equiv \min\left\{\underline{\widehat{ub}}, 1\right\}$ of $\overline{\pi}$.

To conduct inference on π, we seek a critical value r such that:

$$P\left(\underline{ub} \leq \underline{\widehat{ub}} + r\right) = 0.95, \tag{12}$$

as such an r implies

$$\begin{aligned} P\left(\pi \leq \min\left\{\underline{\widehat{ub}} + r, 1\right\}\right) &\geq P\left(\overline{\pi} \leq \min\left\{\underline{\widehat{ub}} + r, 1\right\}\right) \\ &\geq P\left(\underline{ub} \leq \min\left\{\underline{\widehat{ub}} + r, 1\right\}\right) \\ &= P\left(\underline{ub} \leq \underline{\widehat{ub}} + r\right) 1\left[\underline{\widehat{ub}} + r < 1\right] + 1\left[\underline{\widehat{ub}} + r \geq 1\right] \\ &= 0.95 \times 1\left[\underline{\widehat{ub}} + r < 1\right] + 1\left[\underline{\widehat{ub}} + r \geq 1\right] \\ &\geq 0.95 \end{aligned}$$

with the first inequality binding when $\pi = \overline{\pi}$ and the second when $\underline{ub} < \overline{\pi}$.

We can rewrite (12) as:

$$P\left(-\min\left\{\widehat{ub}_1 - \underline{ub}, \widehat{ub}_2 - \underline{ub}, ..., \widehat{ub}_K - \underline{ub}\right\} \leq r\right) = 0.95$$

or equivalently

$$P\left(\max\left\{\underline{ub} - \widehat{ub}_1, \underline{ub} - \widehat{ub}_2, ..., \underline{ub} - \widehat{ub}_K\right\} \leq r\right) = 0.95$$

It is well known that the limiting distribution of $\max\left\{\underline{ub} - \widehat{ub}_1, \underline{ub} - \widehat{ub}_2, ..., \underline{ub} - \widehat{ub}_K\right\}$ depends on which and how many of the upper bound constraints bind. Several approaches to this problem have been proposed which involve conducting pre-tests for which constraints are binding (e.g. Andrews and Barwick, 2012).

We will take an alternative approach to inference that is simple to implement and consistent regardless of the constraints that bind. Our approach is predicated on the observation that:

$$P\left(\max\left\{ub_1 - \widehat{ub}_1, ..., ub_K - \widehat{ub}_K\right\} \leq r\right) \leq P\left(\max\left\{\underline{ub} - \widehat{ub}_1, ..., \underline{ub} - \widehat{ub}_K\right\} \leq r\right) \tag{13}$$

with equality holding in the case where all of the upper bound solutions are identical. We seek an r' such that:

$$P\left(\max\left\{ub_1 - \widehat{ub}_1, ..., ub_K - \widehat{ub}_K\right\} \leq r'\right) = .95 \tag{14}$$

From (13),

$$P\left(\max\left\{\underline{ub} - \widehat{ub}_1, ..., \underline{ub} - \widehat{ub}_K\right\} \leq r'\right) \geq .95$$

with equality holding when all bounds are identical.

A bootstrap estimate $r^* \xrightarrow{p} r'$ of the necessary critical value can be had by considering the bootstrap analog of condition (14) (see Proposition 10.7 of Kosorok, 2008). That is, by computing the 95th percentile of:

$$\max\left\{\widehat{ub}_1 - \widehat{ub}_1^*, ..., \widehat{ub}_K - \widehat{ub}_K^*\right\}$$

across bootstrap replications, where stars refer to bootstrap quantities. An upper limit U of the confidence region for π can then be formed as:

$$U = \min\left\{\widehat{\underline{ub}} + r^*, 1\right\}$$

Note that this procedure is essentially an unstudentized version of the inference method of Chernozhukov et al. (2009) where the set of relevant upper bounds ($\mathcal{V}_0$ in their notation) is taken here to be the set of all upper bounds, thus yielding conservative inference.

We turn now to the lower limit of our confidence interval. Our greatest lower bounds are all of the form:

$$\pi \geq \underline{\pi} \equiv \max\{lb, 0\}.$$

We have the plugin lower bound estimator $\widehat{lb} \xrightarrow{p} lb$. By the same arguments as above we want to search for an r'' such that

$$P\left(lb \geq \widehat{lb} - r''\right) = 0.95.$$

Since $\widehat{lb}$ is just a scalar sample mean, we can choose $r'' = 1.65\sigma_{lb}$ where σ_{lb} is the asymptotic standard error of $\widehat{lb}$ in order to guarantee the above condition holds asymptotically. To account for the propensity score reweighting, we use a bootstrap standard error estimator $\widehat{\sigma}_{lb}$ of σ_{lb} which is consistent via the usual arguments. Thus, our conservative 95% confidence interval for π is:

$$\left[\max\left\{0, \widehat{lb} - 1.65\widehat{\sigma}_{lb}\right\}, \min\left\{\underline{\widehat{ub}} + r^{*}, 1\right\}\right]$$

This confidence interval covers the parameter π with asymptotic probability of at least 95%.

Derivation of Transition Matrix

We prove here that the zero entries of the matrix Π correspond to state transitions that, given our model assumptions, cannot occur in response to the JF experiment.[23] To facilitate our exposition, we start with some definitions:

Definition 1 *An allocation is an earnings and consumption equivalent pair* (E, C)*.*

Definition 2 *For treatment regime* $t \in \{a, j\}$*, the state* $s_t = h\left(\widetilde{E}, D, E^r\right)$ *where the function h maps the triple* $\left(\widetilde{E}, D, E^r\right)$ *from its domain* $\{0,1,2\} \times \{0,1\} \times R^{+}$ *to* $\widetilde{S}$ *as explained in the text.*

Definition 3 *An allocation* (E', C') *is compatible with state* s_t *if* E' *evaluates to the same earnings category as* $\widetilde{E}$ *and: i) if* $D = 0$*,* $C' = E' - \mu_i 1\{E' > 0\}$*, ii) if* $D = 1$ *there exists an* $E^r \in [0, E]$ *such that* $C' = E' - \mu_i 1\{E' > 0\} + G_i^t(E^r) - \phi_i - \eta_i^t 1\{E^r = 0\} - \kappa_i 1\{E' \neq E^r\}$*.*

Definition 4 *An allocation* (E', C') *is available if it is compatible with state* s_t *and* E' *corresponds to one of the earning draws* $\{O_i^1, O_i^2\}$ *or* $E' = 0$*.*

Assumption 1 *Treatment regime a allows a small fixed earning disregard denoted by* δ *and belonging to range 1.*

Proposition 1 *Under either treatment regime, the optimal reporting rule entails either truthful reporting or reporting an earnings level that, without loss of generality, equals* δ*.*

Proof. Omitted. ■

Lemma 1 *Consider transition rate* π_{s_a, s_j} *for any* (s_a, s_j) *such that:*

L.1 $s_j \neq s_a$*;*

L.2 the utility value of any allocation compatible with s_a *is at least as high under regime j as under regime a;*

[23] In future updates we will prove that the nonzero entries of Π correspond to transitions that can occur.

L.3 the utility value of any allocation compatible with s_j is at most as high under regime j as under regime a.

Then, the labor supply model implies that $\pi_{s_a,s_j} = 0$.

Proof. Omitted. ■

Assumption 2 *A woman indifferent between two available allocations will not switch between them in response to a change in treatment regime.*

Lemma 2 *The transition probability π_{s_a,s_j} equals zero for all (s_a, s_j) in the collections:*

$$\left\{ \begin{array}{c} (0n,1n),(0n,2n),(0n,0r),(0n,2u), \\ (1n,0n),(1n,2n),(1n,0r),(1n,2u), \\ (2n,0n),(2n,1n),(2n,0r),(2n,2u), \\ (1u,0n),(1u,1n),(1u,2n),(1u,0r),(1u,2u), \\ (2u,0n),(2u,1n),(2u,2n),(2u,0r) \end{array} \right\} \tag{15}$$

$$\{(1r,0n),(1r,1n),(1r,2n),(1r,0r),(1r,2u)\} \tag{16}$$

$$\{(0r,1n),(0r,1u)\} \tag{17}$$

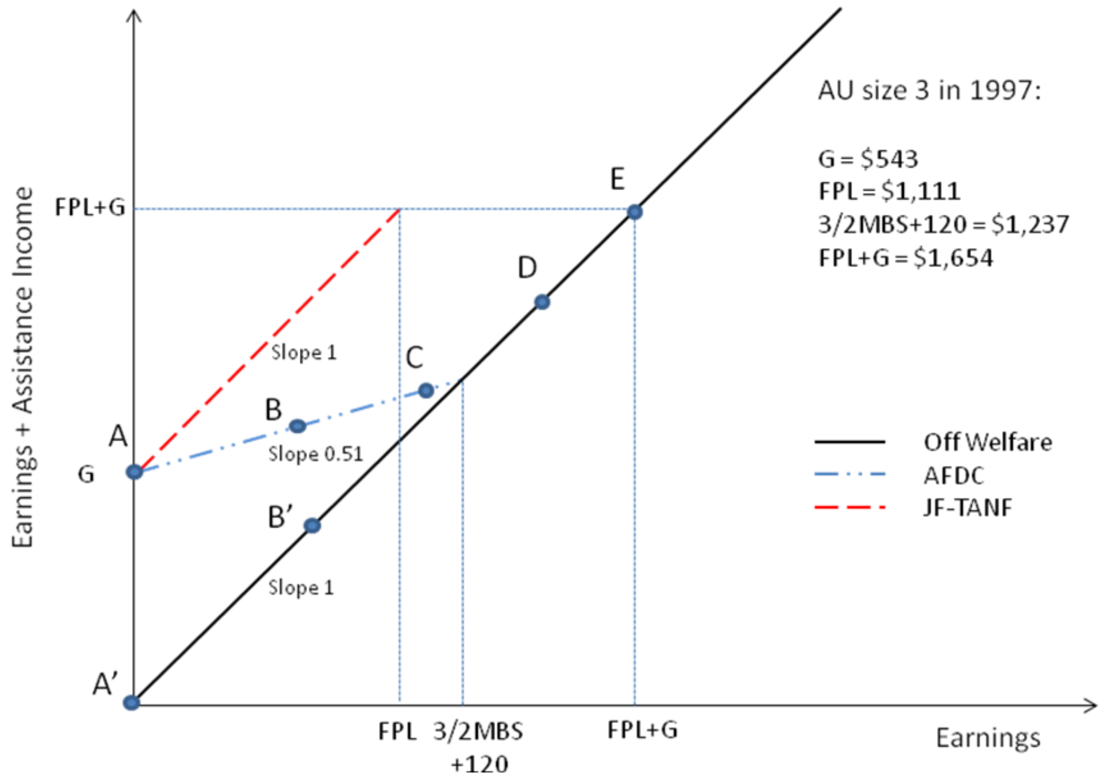
Proof. *We start with the collection in (15). It is sufficient to prove that these transitions obey conditions L.1-L.3. L.1 holds trivially. L.2 and L.3 are met because all allocations compatible with states $(0n, 1n, 2n, 1u, 2u)$ have the same value under either regime.*

Turning to the collection in (16), allocations compatible with state $1r$ must yield utility at least as high under JF as under AFDC because $G_i^a(E) < \overline{G}_i$ for all $E \leq FPL_i$. Therefore condition L.2 is met. L.3 follows because $(0n, 1n, 2n, 1u, 2u)$ have the same value under either regime.

Next, we turn to the collection in (17). Here we argue by contradiction that $\pi_{0r,1n}$ and $\pi_{0r,1u}$ must be zero. Suppose first that there is a woman i who optimally chooses an allocation compatible with state $0r$ under regime a and switches to an available allocation compatible with state $1n$ under regime j, say at O_i^k. By optimality of the allocation compatible with state $0r$ under regime a we have $U_i\left(0, \overline{G}_i - \phi_i - \eta_i^a\right) \geq U_i(0,0)$ which implies $\overline{G}_i - \phi_i \geq \eta_i^a$ where $\eta_i^a \geq 0$. By optimality of the allocation compatible with state $1n$ under regime j, $U_i\left(O_i^k, O_i^k - \mu_i\right) \geq U_i\left(O_i^k, O_i^k - \mu_i + \overline{G}_i - \phi_i\right)$ which implies $\overline{G}_i - \phi_i \leq 0$. If $\eta_i^a > 0$ or $\left(\eta_i^a = 0, \overline{G}_i \neq \phi_i\right)$ this entails an immediate contradiction. If $\left(\eta_i^a = 0, \overline{G}_i = \phi_i\right)$ then the woman must have been indifferent between the allocation compatible with state $0p$ and that *compatible with* $0n$ *under regime* a *which means that* $U_i(0,0) \geq U_i\left(O_i^l, O_i^l - \mu_i\right)$ *for any* O_i^l *in range 1 including* O_i^k*. By optimality of the allocation compatible with state* $1n$ *under* j*,* $U_i\left(O_i^k, O_i^k - \mu_i\right) \geq U_i(0,0)$*. Thus, if this last inequality is strict we have a contradiction. If an equality obtains, the woman must have been indifferent under regime* a *between the allocation compatible with state* $0n$ *and the allocation entailing earnings* O_i^k *off assistance. From assumption 2, if she did not choose* O_i^k *over* $0n$ *under regime* a*, given indifference, she will not make a different choice under* j*.*

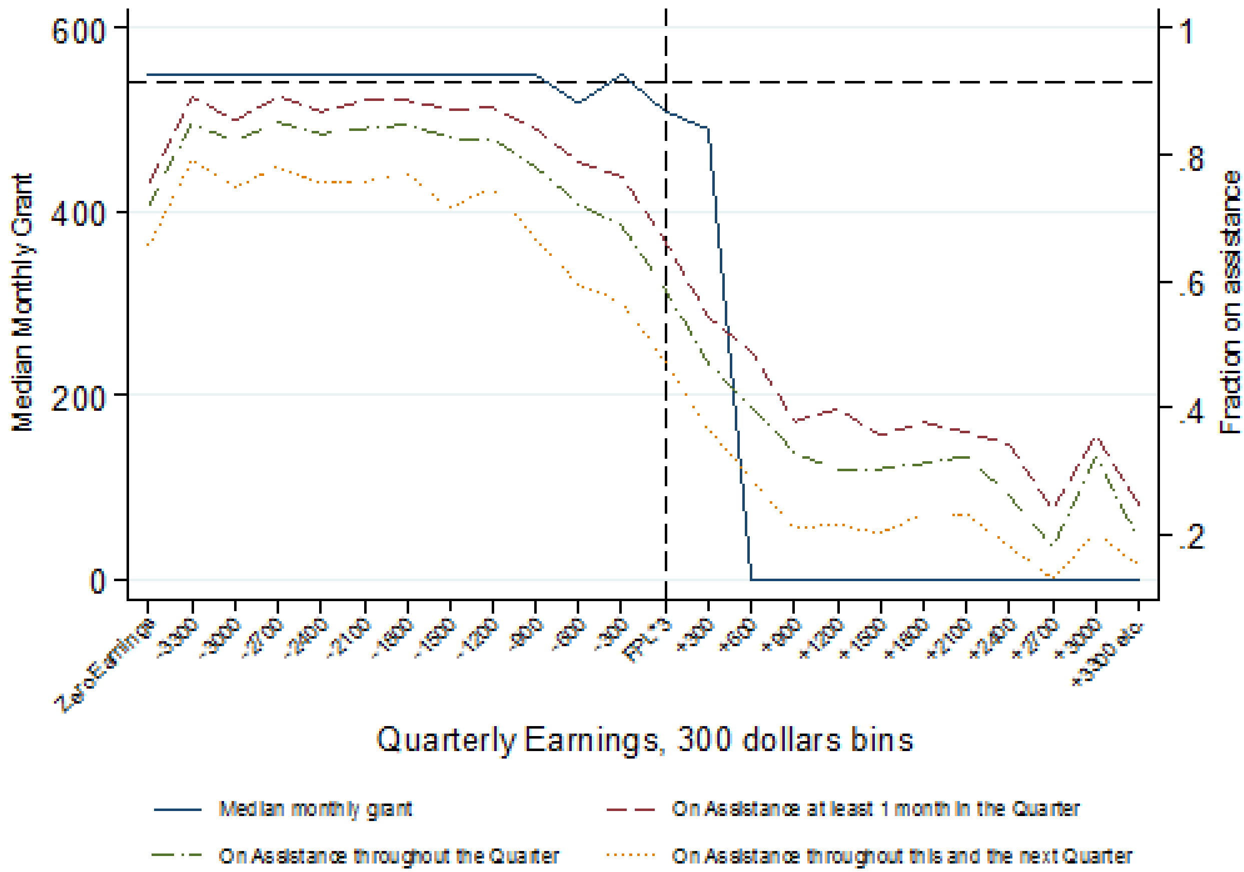
Next, suppose that there is a woman i *who optimally chooses an allocation compatible with state* $0r$ *under* a *and switches to an available allocation compatible with state* $1u$ *under* j*. By Proposition 1, no allocation compatible with* $1u$ *may be optimal under regime* j *which produces a contradiction.* ■

Figure 1: Hypothetical Budget Sets under AFDC and JF



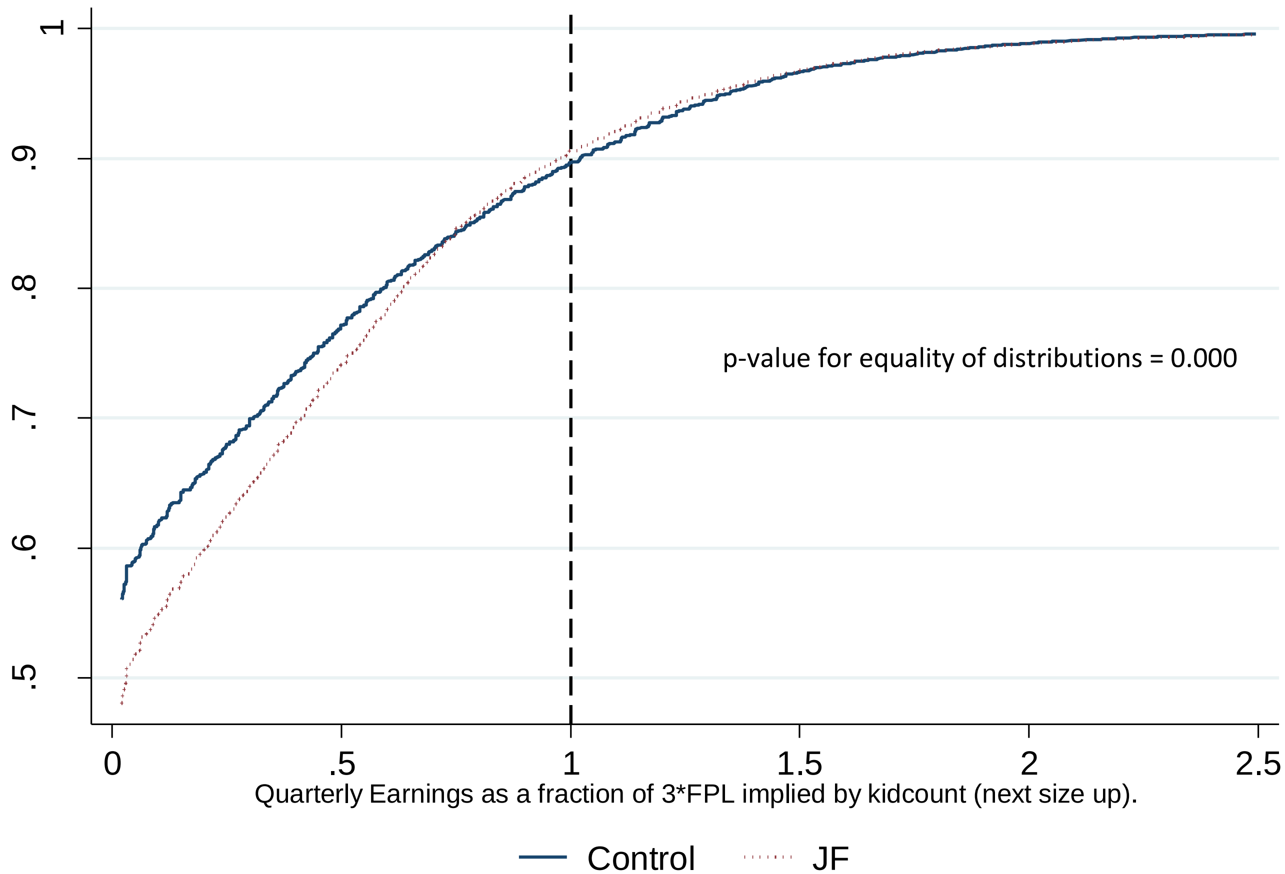
Notes: Figure provides hypothetical monthly budget constraints faced by assistance unit of size 3 under AFDC and JF policy rules. Figure assumes household would have access to the \$30 + 1/3rd disregard under AFDC. FPL refers to federal poverty line, G is base grant amount, and MBS is AFDC needs standard.

Figure 2: Benefits as a Function of Earnings in JF Sample



Notes: The sample includes the JF-TANF cases, all months in Q1–Q7 post-RA such that imputed AU size is between 2 and 5. The horizontal line is drawn at (unrounded) maximum grant of a size 3 AU in 1997. Grant amounts and earnings are rescaled using, respectively, the maximum grant and FPL of size 3 AU in 1997. Median monthly grant computed across all cases including those not on welfare in the quarter.

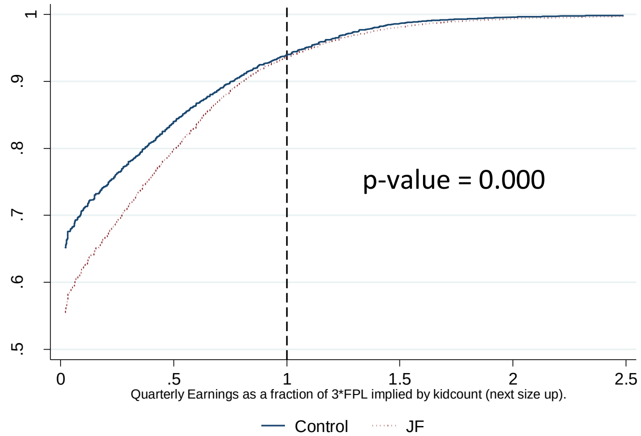
Figure 3: CDFs of Quarterly Earnings Relative to 3 x Monthly Federal Poverty Line



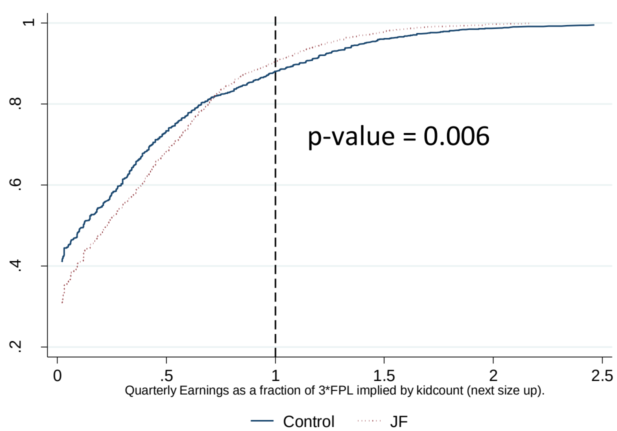
Notes: Figure gives reweighted CDFs of quarterly UI earnings (in quarters 1-7 post-RA) in JF and AFDC samples relative to 3 x the monthly federal poverty line associated with year and AU size. AU size determined by baseline survey question kidcount. To deal with increases in family size since random assignment, we use next AU size up relative to size directly implied by kidcount (see text for details). The bootstrapped p-value for a Kolmogorov-Smirnov test of equality of the two distributions is 0.000 (based on 1,000 replications, see Appendix for details).

Figure 4: CDFs of Quarterly Earnings Relative to 3 x Federal Poverty Line (Subgroups)

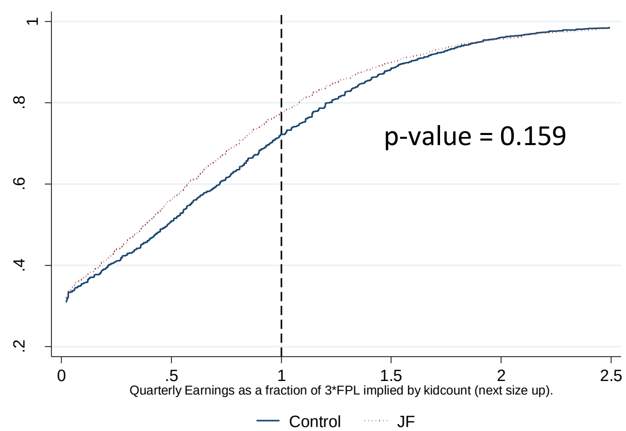
a) Zero Earnings pre-RA



b) Low Earnings pre-RA



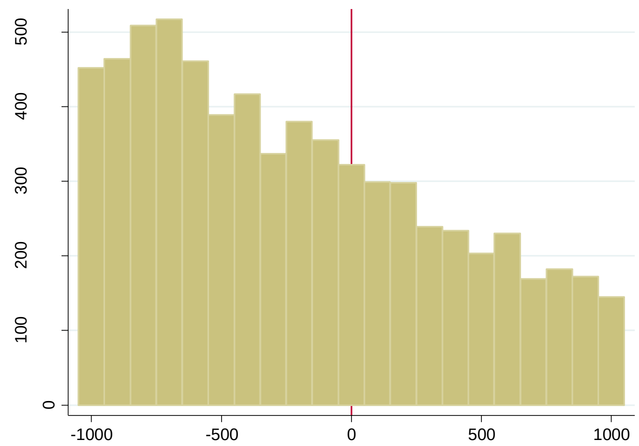
c) High Earnings pre-RA



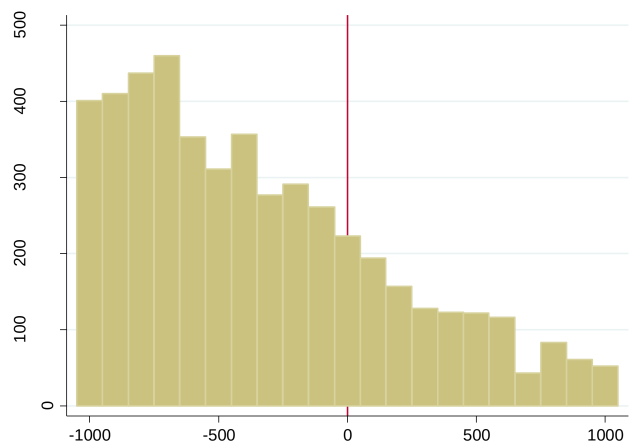
Notes: Figures give reweighted CDFs of quarterly UI earnings (in quarters 1-7 post-RA) in JF and AFDC samples relative to 3 x the monthly federal poverty line associated with year and AU size. AU size determined by baseline survey question kidcount. To deal with increases in family size since random assignment, we use next AU size up relative to size directly implied by kidcount (see text for details). The bootstrapped p-value for a Kolmogorov-Smirnov test of equality of the two distributions is is 0.000 for the Zero Earning subgroup, 0.006 for the Low Earning subgroup and 0.159 for the High Earning subgroup (based on 1,000 replications, see Appendix for details).

Figure 5: Distribution of Quarterly Earnings Centered at 3 x Monthly Federal Poverty Line

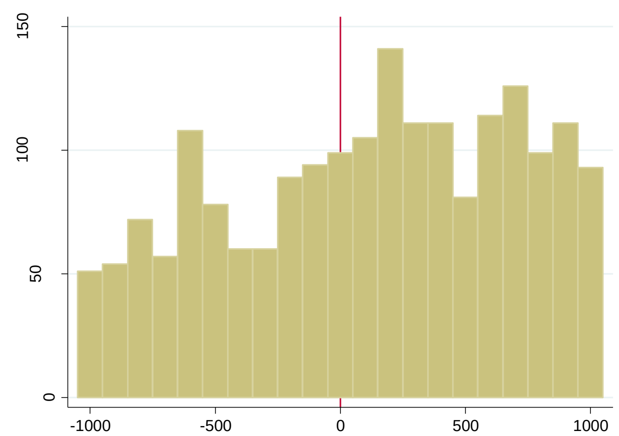
a) Unconditional



b) On Assistance all 3 Months of the Quarter



c) Off Assistance all 3 Months of the Quarter



Notes: Restricted to Jobs First sample quarters 1-7 post-RA. Assistance unit size has been inferred from monthly aid payment. AU sizes above 8 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 size and the calendar year of the quarterly observation.

Figure 6: Earnings and Participation Choices with Under-reporting

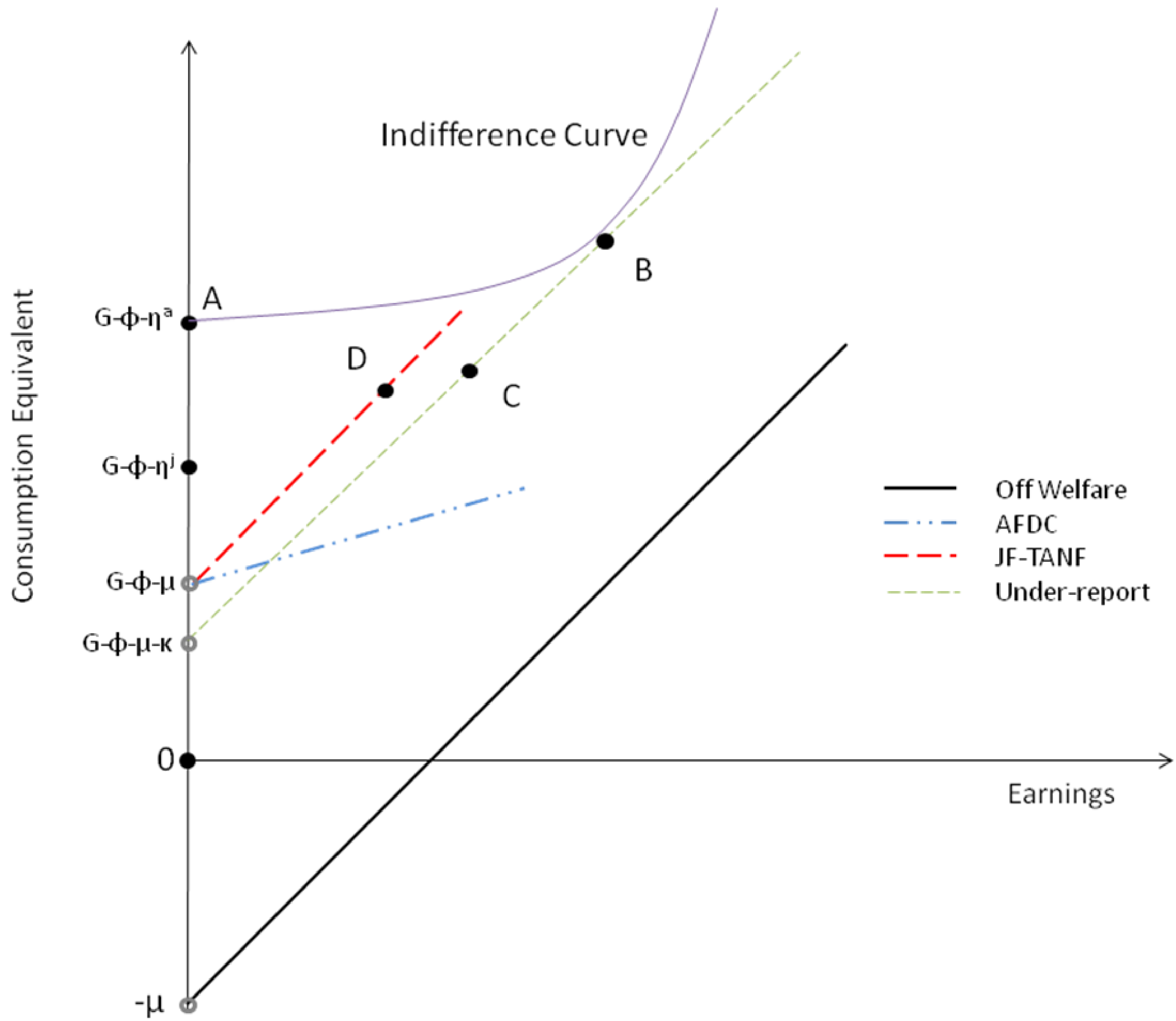


Table 1 - Summary of Policy Differences Between AFDC and Jobs First

	JF	AFDC
Eligibility	Earnings Below Poverty Line	Earnings level at which benefits are exhausted (see disregard parameters below)
Earnings disregard	All earned income disregarded up to poverty line	Months 1-4: \$120 + 1/3 Months 5-12: \$120 Month >12: \$90
Time Limit	21 months (6 month extensions possible)	None
Work requirements	Mandatory work first (exempt if child <1)	Education / training (exempt if child < 2)
Other	●Sanctions (moderate enforcement) ●Asset limit \$3,000 ●Partial family cap (50 percent) ●Two years transitional Medicaid ●Child care assistance ●Child support: \$100 disregard, full pass-through	●Sanctions (rarely enforced) ●Asset limit \$1,000 ●100 hour rule and work history requirement for two-parent families ●One year transitional Medicaid ●Child support: \$50 disregard, \$50 maximum pass-through

Sources: Adams-Ciardullo et al. (2002) and Bitler, Gelbach, and Hoynes (2005).

Table 2: Mean Sample Characteristics

	Overall Sample				Zero Earnings Q7 pre-RA				Low Earnings Q7 pre-RA				High Earnings Q7 pre-RA			
	Jobs First	AFDC	Difference	Difference (adjusted)	Jobs First	AFDC	Difference	Difference (adjusted)	Jobs First	AFDC	Difference	Difference (adjusted)	Jobs First	AFDC	Difference	Difference (adjusted)
Demographic Characteristics																
White	0.362	0.348	0.014	0.001	0.331	0.321	0.010	-0.002	0.443	0.421	0.022	0.000	0.425	0.385	0.040	0.003
Black	0.368	0.371	-0.003	0.000	0.360	0.349	0.011	0.001	0.367	0.428	-0.061	-0.002	0.406	0.403	0.003	-0.004
Hispanic	0.207	0.216	-0.009	-0.001	0.250	0.267	-0.016	0.000	0.126	0.098	0.028	-0.001	0.086	0.127	-0.041	-0.004
Never married	0.624	0.631	-0.007	0.000	0.631	0.627	0.005	0.000	0.664	0.708	-0.044	0.004	0.550	0.568	-0.019	-0.002
Div/wid/sep/living apart	0.317	0.312	0.005	0.000	0.314	0.320	-0.006	0.000	0.272	0.237	0.035	-0.008	0.376	0.357	0.019	0.003
HS dropout	0.331	0.313	0.018	0.000	0.369	0.370	-0.001	0.000	0.303	0.224	0.078	-0.001	0.180	0.165	0.014	0.007
HS diploma/GED	0.550	0.566	-0.016	0.001	0.521	0.521	0.000	0.000	0.583	0.668	-0.085	0.000	0.655	0.651	0.004	-0.014
More than HS diploma	0.066	0.062	0.004	0.000	0.060	0.051	0.009	0.009	0.054	0.056	-0.002	-0.001	0.109	0.115	-0.006	0.017
More than 2 Children	0.227	0.206	0.021	0.000	0.252	0.242	0.010	0.000	0.179	0.123	0.056	-0.007	0.157	0.142	0.015	-0.004
Mother younger than 25	0.289	0.297	-0.007	0.000	0.288	0.269	0.019	-0.001	0.381	0.448	-0.067	0.007	0.204	0.256	-0.051	-0.002
Mother age 25-34	0.410	0.418	-0.007	0.000	0.411	0.418	-0.007	0.000	0.361	0.358	0.004	-0.001	0.456	0.478	-0.022	-0.001
Mother older than 34	0.301	0.286	0.015	0.000	0.301	0.313	-0.012	0.001	0.258	0.194	0.064	-0.007	0.340	0.266	0.074	0.003
Average quarterly pretreatment values																
Earnings	679	786	-107***	-1	175	189	-14	2	770	865	-94	-7	2924	3209	-285	-23
	(1304)	(1545)	(41)	(51)	(465)	(490)	(17)	(4)	(683)	(710)	(51)	(28)	(1890)	(2441)	(159)	(443)
Cash welfare	891	835	56**	-1	1041	1013	28	0	785	696	90	1	298	232	66	-24
	(806)	(785)	(23)	(2)	(812)	(802)	(28)	(3)	(722)	(660)	(51)	(22)	(517)	(417)	(34)	(80)
Food stamps	352	339	13	0	397	394	2	1	329	297	32	0	168	151	17	-14
	(320)	(304)	(9)	(1)	(326)	(310)	(11)	(1)	(300)	(267)	(21)	(20)	(232)	(219)	(16)	(45)
Fraction of pretreatment quarters with																
Any earnings	0.322	0.351	-0.029***	0.000	0.137	0.144	-0.008	0.000	0.667	0.694	-0.027	0.000	0.839	0.864	-0.025	0.003
	(0.363)	(0.372)	(0.011)	(0.001)	(0.211)	(0.216)	(0.007)	(0.001)	(0.273)	(0.259)	(0.019)	(0.008)	(0.216)	(0.181)	(0.015)	(0.02)
Any cash welfare	0.573	0.544	0.029**	-0.001	0.644	0.630	0.014	0.000	0.564	0.532	0.032	0.001	0.254	0.198	0.056	-0.010
	(0.452)	(0.450)	(0.013)	(0.001)	(0.440)	(0.442)	(0.015)	(0.001)	(0.448)	(0.441)	(0.032)	(0.01)	(0.365)	(0.305)	(0.025)	(0.04)
Any food stamps	0.607	0.598	0.009	0.000	0.664	0.669	-0.004	0.001	0.602	0.588	0.014	0.002	0.345	0.312	0.033	-0.009
	(0.438)	(0.433)	(0.013)	(0.001)	(0.428)	(0.422)	(0.015)	(0.001)	(0.433)	(0.423)	(0.031)	(0.01)	(0.391)	(0.362)	(0.028)	(0.04)
# of Cases	2,396	2,407			1,677	1,623			357	397			362	387		

Notes: Adjusted differences are computed via propensity score reweighting. Numbers in parentheses are standard deviations for columns 1, 2, 3, 5, 6, 7; those for columns 4 and 8 are standard errors calculated via 1,000 bootstrap replications.

Table 3: Mean Outcomes Post-Random Assignment

	Overall			Zero Earnings Q7 pre-RA			Low Earnings Q7 pre-RA			High Earnings Q7 pre-RA		
	Jobs First	AFDC	Adjusted Difference	Jobs First	AFDC	Adjusted Difference	Jobs First	AFDC	Adjusted Difference	Jobs First	AFDC	Adjusted Difference
Average Earnings	1,191	1,086	**105**	930	751	**179**	1,362	1,291	**70**	2,124	2,362	**-238**
	(29)	(30)	**(36)**	(32)	(30)	**(42)**	(66)	(101)	**(118)**	(114)	(151)	**(179)**
Fraction of quarters	0.520	0.440	**0.080**	0.445	0.349	**0.096**	0.691	0.590	**0.100**	0.680	0.690	**-0.011**
with positive earnings	(0.008)	(0.007)	**(0.010)**	(0.009)	(0.009)	**(0.012)**	(0.018)	(0.022)	**(0.028)**	(0.022)	(0.028)	**(0.035)**
Fraction of quarters with earnings below	0.665	0.710	**-0.046**	0.731	0.789	**-0.058**	0.570	0.636	**-0.066**	0.477	0.438	**0.039**
monthly FPL (AU size implied by kidcount+1)	(0.004)	(0.005)	**(0.006)**	(0.004)	(0.004)	**(0.006)**	(0.012)	(0.017)	**(0.020)**	(0.019)	(0.025)	**(0.030)**
Fraction of quarters with earnings below	0.906	0.897	**0.009**	0.938	0.940	**-0.002**	0.906	0.881	**0.025**	0.777	0.722	**0.055**
3FPL (AU size implied by kidcount+1)	(0.007)	(0.007)	**(0.009)**	(0.008)	(0.008)	**(0.010)**	(0.019)	(0.021)	**(0.027)**	(0.023)	(0.030)	**(0.036)**
Fraction of quarters on welfare	0.748	0.674	**0.074**	0.771	0.718	**0.053**	0.764	0.674	**0.091**	0.637	0.475	**0.162**
	(0.007)	(0.007)	**(0.010)**	(0.008)	(0.008)	**(0.011)**	(0.019)	(0.022)	**(0.029)**	(0.024)	(0.033)	**(0.040)**
Average earnings in quarters	929	526	**403**	762	404	**359**	1,123	694	**428**	1,524	1,075	**449**
with any month on welfare	(24)	(19)	**(28)**	(25)	(18)	**(30)**	(53)	(48)	**(69)**	(96)	(119)	**(147)**
Fraction of quarters with no earnings and	0.363	0.437	**-0.074**	0.426	0.508	**-0.082**	0.231	0.334	**-0.103**	0.221	0.220	**0.001**
at least one month on welfare	(0.007)	(0.007)	**(0.010)**	(0.009)	(0.009)	**(0.012)**	(0.015)	(0.022)	**(0.027)**	(0.019)	(0.028)	**(0.033)**
# of cases	2,318	2,324		1,630	1,574		343	384		345	366	

Notes: Sample covers quarters 1-7 post-random assignment. Adjusted differences are computed via propensity score reweighting. Numbers in parentheses are standard errors calculated via 1,000 bootstrap replications. Sample units with kidcount missing are excluded.

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**
				(0.006)	(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)	**(0.012)**
Pr(State=1p)	0.342	0.185	0.157	0.343	0.184	**0.159**
				(0.008)	(0.006)	**(0.009)**
Pr(State=2p)	0.022	0.009	0.013	0.023	0.009	**0.014**
				(0.002)	(0.001)	**(0.002)**
# of 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. 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. The letter n indicates welfare nonparticipation throughout the quarter while the letter p indicates welfare participation throughout the quarter. Poverty line computed under assumption AU size is one greater than amount implied by baseline kidcount variable. Adjusted probabilities are adjusted via the propensity score reweighting algorithm described in the Appendix. Standard errors computed using 1,000 bootstrap replications.

Table 5: Point and Set-identified Transition Probabilities

	Estimate	Standard Error	95% CI (deterministic bound)	95% CI (conservative)
Point-identified				
Off welfare, low earnings → On welfare, low earnings ($\pi_{1n,1r}$)	0.381	0.038	[0.306, 0.455]	[0.306, 0.455]
Set-identified				
On welfare, not working → Off welfare, not working ($\pi_{0r,0n}$)	{0.000, 0.169}		[0.000, 0.210]	[0.000, 0.245]
On welfare, not working → On welfare, low earnings ($\pi_{0r,1r}$)	{0.000, 0.169}		[0.000, 0.210]	[0.000, 0.250]
On welfare, not working → Off welfare, high earnings ($\pi_{0r,2n}$)	{0.000, 0.154}		[0.000, 0.170]	[0.000, 0.302]
On welfare, not working → On welfare, high earnings ($\pi_{0r,2u}$)	{0.031, 0.051}		[0.022, 0.059]	[0.022, 0.131]
Off welfare, not working → On welfare, low earnings ($\pi_{0n,1r}$)	{0.059, 0.618}		[0.000, 0.755]	[0.000, 0.875]
Off welfare, high earnings → On welfare, low earnings ($\pi_{2n,1r}$)	{0.281, 1.000}		[0.193, 1.000]	[0.193, 1.000]
On welfare, high earnings → Off welfare, low earnings ($\pi_{2u,1r}$)	{0.000, 1.000}		[0.000, 1.000]	[0.000, 1.000]
Composite Margins				
Not working → Working ($\pi_{0,1+}$)	0.168	0.020	[0.129, 0.206]	[0.129, 0.206]
Off welfare → On welfare ($\pi_{n,p}$)	{0.232, 0.444}		[0.188, 0.486]	[0.188, 0.509]
On welfare → Off welfare ($\pi_{p,n}$)	{0.000, 0.118}		[0.000, 0.147]	[0.000, 0.173]
On welfare, not working → Off welfare ($\pi_{0r,n}$)	{0.000, 0.169}		[0.000, 0.210]	[0.000, 0.244]

Notes: Estimates inferred from probabilities in Table 4, see text for formulas. Low earnings refers to quarterly earnings less than or equal to three times the monthly federal poverty line (E=1), high earnings refer to quarterly earnings above three times the monthly federal poverty line (E=2). Numbers in braces are estimated upper and lower bounds, numbers in brackets are 95% confidence intervals. Column labeled "deterministic bound" ignores uncertainty in which moment inequalities bind. Column labeled "conservative" uses inference procedure described in appendix which covers true parameter with at least 95% probability regardless of which constraints bind. See text for details.

Table A1: Cross Tabulation of grant-inferred AU size and kidcount

	kidcount				
	0	1	2	3	Total
Inferred Size					
1	0.17	0.08	0.04	0.01	0.05
2	0.53	0.84	0.19	0.06	0.42
3	0.17	0.06	0.72	0.17	0.29
4	0.11	0.01	0.05	0.53	0.17
5	0.00	0.00	0.00	0.14	0.04
6	0.00	0.00	0.00	0.01	0.00
7	0.03	0.00	0.00	0.07	0.02
8	0.00	0.00	0.00	0.00	0.00
# of monthly observations	840	11,361	8,463	8,043	28,707

Notes: Analysis conducted on Jobs First sample over quarters 1-7 post-random assignment. Kidcount variable, which is obtained from baseline survey, is tabulated conditional on non-missing. The AU size is inferred from (rounded) monthly grant amounts and MAP values up to a size of 8. Starting with size 5, the unique correspondence between AU size and rounded grant amount obtains only for units which do not receive housing subsidies. Thus, inferred sizes from 5 to 8 rest on this assumption. The size inferred during months on assistance is imputed forward to months off assistance and to months that otherwise lack an inferred size.