
Jonah B. Gelbach
University of Pennsylvania Carey Law School

Follow this and additional works at: https://scholarship.law.upenn.edu/faculty_scholarship

Part of the Labor and Employment Law Commons, Labor Economics Commons, Law and Economics Commons, Policy History, Theory, and Methods Commons, Political Economy Commons, Politics and Social Change Commons, Public Economics Commons, Public Law and Legal Theory Commons, Public Policy Commons, Social Welfare Commons, Social Welfare Law Commons, and the State and Local Government Law Commons

Repository Citation
https://scholarship.law.upenn.edu/faculty_scholarship/1671

This Article is brought to you for free and open access by Penn Law: Legal Scholarship Repository. It has been accepted for inclusion in Faculty Scholarship at Penn Law by an authorized administrator of Penn Law: Legal Scholarship Repository. For more information, please contact PennlawIR@law.upenn.edu.
Uncontrolled Experiments From the Laboratories of Democracy: Traditional Cash Welfare, Federalism, and Welfare Reform

Draft: May 5, 2016

Jonah B. Gelbach¹

University of Pennsylvania Law School

¹ This chapter was substantially completed in 2012, and I have not sought to update it substantively since then. I am grateful to numerous coauthors and colleagues for many conversations over the years concerning welfare policy and empirical research. Any mistakes that remain are of course my own.
It is one of the happy incidents of the federal system that a single courageous state may, if its citizens choose, serve as a laboratory; and try novel social and economic experiments without risk to the rest of the country.”


I. Introduction

The story of the rise and fall of traditional cash welfare in the United States is complex in some ways, but straightforward in others. For roughly six decades, the primary source of federal cash assistance to the non-elderly, non-disabled poor was the Aid to Families with Dependent Children (AFDC) program. While the details of the program’s rules could be technical and complex, AFDC was basically a check-cutting program. It was one intended to do little more than its name suggested—provide assistance to (low-income) families with dependent children. For its first three decades, the AFDC program forged a relatively low profile. In the mid-1960s, though, benefit levels began to rise, perhaps as a result of a structurally expanded federal role in AFDC funding. Caseloads rose, too. And perhaps as a result, the seeds of a political backlash were planted.

By the period between the late 1980s and the mid-1990s, many policy makers at both the state and federal levels were chafing at program’s relatively rigid rules. The “laboratories of democracy” concept proved useful in organizing opposition to the existing system, with states like Wisconsin and Michigan asking for and getting waivers—permission to use federal funds to operate programs with important structural differences from the AFDC program. Bill Clinton’s election in 1992 brought more openness to such changes among Democrats, and in 1996 President Clinton made good on his earlier campaign promise to “end welfare as we know it”2 by signing the Personal Responsibility and Work Opportunity Reconciliation Act (PRWORA),3 which eliminated the AFDC program formally. In its place is the Temporary Assistance for Needy Children (TANF) program, the block-grant umbrella covering a patchwork of state programs.

News accounts and political (and policy) discussions brim with what can be called the triumphalist consensus concerning welfare reform. For example, read current Brookings scholar Ron Haskins, who played an important role in shepherding PRWORA through Congress in his role as a congressional staff member. Toward the end of his ten-years-thence account of the fight to enact PRWORA, in a chapter subtitled “The Triumph of Work,” Haskins writes that “[s]weeping reforms have produced sweeping effects.”4 One would expect to see this claim followed with a list of some of these sweeping “effects” that welfare reform has “produced”. But what comes in Haskins’s next sentence is not a causal claim at all. Rather: “The evidence that welfare use has dropped, that work and earnings have increased and that child poverty has fallen is overwhelming.”5

---

3 Public Law 104-193; see 110 Stat. 2105 for text.
5 Haskins (2006), at 362 (my emphasis). I don’t mean to pick on Haskins, who is a serious and fair-minded scholar. One can find many declarations of victory from across the political spectrum. See, for example, virtually anything.
A moment’s reflection reveals that overwhelming evidence of a thing’s occurrence is not even trivial evidence that the thing has happened for any particular reason. As the discussion below suggests, the evidence for the triumphalist consensus largely boils down to a claim of post-welfare-reform-ergo-propter-welfare-reform. Long before the reforms of the 1990s, scholars knew that there was little empirical evidence that AFDC’s incentive effects were all that substantial. And so it should not be much of a surprise that eliminating AFDC does not seem to have caused very much of the major changes in welfare use and labor market outcomes at the root of post-hoc triumphalism. Welfare reform did occur against the backdrop of a booming labor market, and it also occurred at a time when social stigma directed at welfare use seems to have greatly increased; perhaps these are alternative explanations for the trends observed after welfare reform. It will be worth keeping such possibilities in mind through the balance of this chapter.

In Part II of this chapter, I lay out the basic structure of the old AFDC program and the ways in which the state and federal reforms of the 1990s transformed federally funded cash assistance. I will not attempt a comprehensive review of the vast micro-econometric literature on the incentive effects of both the AFDC program and its post-reform successors. But where appropriate, I discuss general conclusions and some specific findings on the behavioral effects of both the old and new programs. While empirical evidence certainly suggests the AFDC program had incentive effects, it hardly justifies the extremity of the negative views held by many. The program did not eliminate poverty, to be sure. But given its basic structure and funding levels, it is unreasonable to have expected it to. And while the increase in employment of low-skilled single mothers following welfare reform surprised many observers, the common refrain that welfare reform “worked” is built on notably shaky empirical ground, a point revisited in Part IV.

In Part III, I evaluate welfare reform from the perspective of Justice Brandeis’s laboratories metaphor, taking an atypically jaded view of the devolution of the 1990s. Given the vast array of differences in state welfare plans, the reforms of the 1990s might best be thought of as uncontrolled experimentation, undermining the Brandeisian notion of states as useful laboratories.

In Part IV, I present some empirical evidence from Connecticut’s Jobs First welfare reform plan, which was initially implemented as an AFDC waiver and subsequently formed the

---


core of Connecticut’s post-PRWORA welfare system. This evidence suggests that the nuts and bolts of welfare reform—actual programmatic differences—had surprisingly little to do with the late-1990s increase in employment among the disadvantaged. Moreover, Connecticut’s welfare reform reduced income for at least some women, which implies that one’s assessment of whether welfare reform “worked” is unavoidably normative. While I do not imagine that policy makers are likely to reintroduce AFDC, I believe that evidence from Connecticut (and similar evidence from other states)—together with the absence of any evidence that welfare reform explains much of the drop in caseloads or increase in employment observed over the periods researchers have studied—calls for a reassessment of just how important welfare program details were in driving the changes of the late 1990s and early 2000s.7

II. A Brief History of National Cash Welfare Policy

Section 401 of the Social Security Act of 1935 established the Aid to Dependent Children (ADC) program “[f]or the purpose of enabling each State to furnish financial assistance, as far as practicable under the conditions in such State, to needy dependent children.”8 Sixty years later, President Clinton signed the Republican Congress’s Personal Responsibility and Work Responsibility Act of 1996, eliminating ADC’s successor program, AFDC, and replacing it with Temporary Assistance to Needy Families (TANF). PRWORA’s clearly stated purposes departed substantially from the check-cutting objective of 1935’s ADC program. Subsection (a) of 42 U.S.C.A. § 601 declares that

The purpose of this part is to increase the flexibility of States in operating a program designed to--

(1) provide assistance to needy families so that children may be cared for in their own homes or in the homes of relatives;
(2) end the dependence of needy parents on government benefits by promoting job preparation, work, and marriage;
(3) prevent and reduce the incidence of out-of-wedlock pregnancies and establish annual numerical goals for preventing and reducing the incidence of these pregnancies; and
(4) encourage the formation and maintenance of two-parent families.

Thus, a program that started out with the simple purpose of delivering cash assistance became one with a mandate not only to provide that assistance—temporarily—but also to: end dependence, whether through work, job “preparation” or marriage; affect pregnancy; affect marriage; and affect household formation and structure. Lest anyone doubt PRWORA’s decisive shift away from assistance as a matter of right, subsection (b) of 42 U.S.C.A. §601 trumpets that “This part shall not be interpreted to entitle any individual or family to assistance under any State

7 That leaves the important question of what does explain these changes. The booming late-1990s labor market is one explanation, though the boom did end. Another possible explanation is hard-to-quantify cultural changes. Sorting out the relative contributions of these factors, if it is even possible, is far beyond the scope of this review.
program funded under this part.”9 Relatedly, it is no accident that TANF is not ANF: we will see that the “Temporary” qualifier does real work.

In section II.A, I begin by discussing the basics of the AFDC program, including a bit of the considerable programmatic history that transpired between the mid-1930s and the mid-1990s.10 I also discuss empirical evidence on some important aspects of AFDC’s behavioral effects. I offer some bigger-picture thoughts concerning AFDC in section II.B. In section II.C, I discuss some of the key reforms that state waivers and PRWORA introduced, relative to AFDC, and I also discuss empirical evidence on welfare reform’s effects.

A. The AFDC Program’s Basic Features

1. The Individual Level

At the individual level, the AFDC program can be thought of as a negative income tax (NIT) program with a categorical eligibility requirement.11 I discuss the NIT part first and then turn to the categorical eligibility requirement.

a. NIT Aspects of AFDC

The simplest NIT program involves two parameters: a maximum grant amount, $G_{AFDC}$, and a benefit reduction rate, $t_{AFDC}$. A person with pre-tax-and-transfer earnings $Y$ who faces this two-parameter NIT program will either receive a transfer payment from the government, or pay a tax to the government according to whether the quantity

$$P \equiv G_{AFDC} - t_{AFDC}Y$$

is positive (transfer received) or negative (tax paid). It is conventional to use a name like “$t$” for the benefit reduction rate because it functions like an income tax by reducing the transfer payment $P$ by $t_{AFDC}$ dollars for every dollar of pre-transfer income earned. As such, a greater value of $t_{AFDC}$ has all the same work disincentive effects of an income tax.

To visualize the incentive effects of an NIT program, consider panels (a) and (b) of Figure 1. The straight line in panel (a) shows the tradeoff between hours of leisure, which are depicted on the horizontal axis, and post-tax-and-transfer consumption, which is depicted on the vertical axis, when there is no transfer program. For simplicity, I assume that there is a constant marginal tax rate at all possible earnings levels in this panel, so that its budget line has constant slope.

In panel (b), the solid line represents the budget line under an NIT program with maximum benefit level $G_{NIT}$ and benefit reduction rate $t_{NIT}$. To the left of $H^{\text{Breakeven}}$, the person

---

10 This discussion is not meant to be exhaustive; I mean only to emphasize only some of the aspects of this history that seem most important from today’s vantage point.
earns more than $y^{\text{Breakeven}} \equiv \frac{g_{\text{NIT}}}{t_{\text{NIT}}}$ and so she pays a positive tax to the government. To the right of $H^{\text{Breakeven}}$, the opposite is true, and so the worker receives a net transfer from the government. The amount of any tax or transfer is given by the height separating the solid and dashed lines.

**Figure 1: Simplified Budget Sets with No Transfer Program, with an NIT, and with AFDC**

Leaving aside categorical eligibility considerations to be discussed below, the AFDC program can be thought of as a modified version of the simple NIT program depicted in panel (b) of Figure 1. The first modification is that the AFDC program never actually required anyone to
pay a tax back to the government. Algebraically, this means the AFDC program involved a transfer payment equal to

\[ P^* \equiv \max [G_{\text{AFDC}} - t_{\text{AFDC}}Y, 0]. \]

Such a program transfers assistance to those with sufficiently low income—below the break-even level \( Y^{\text{Breakeven}} \equiv \frac{G_{\text{AFDC}}}{t_{\text{AFDC}}} \). Panel (c) of Figure 1 shows that the AFDC program results in the budget line given by the solid line when leisure hours are between \( H^{\text{Max}} \) and \( H^{\text{Breakeven}} \), stapled onto the dashed line when leisure hours are less than \( H^{\text{Breakeven}} \). The result is just the NIT program from panel (b) without the part of the NIT budget line to the left of \( H^{\text{Breakeven}} \).

Notice that for a fixed value of \( t_{\text{AFDC}} \), increases in the maximum grant \( G_{\text{AFDC}} \) cause \( Y^{\text{Breakeven}} \) to rise. The more generous is the maximum grant level, the greater the share of people who will be eligible for the program—and so also the share of people who will face work disincentives from the implicit tax \( t_{\text{AFDC}} \). This effect is known as the “mechanical effect” on eligibility.

Increases in \( t_{\text{AFDC}} \) reduce \( Y^{\text{Breakeven}} \) and thus have the opposite effect on the share of people who will be program-eligible. While this reduces the mechanical effect, increases in \( t_{\text{AFDC}} \) also lead to greater work disincentives at the margin for those people who receive AFDC. This is a first type of behavioral effect of NIT programs: at the margin, they distort what labor economists think of as the labor-leisure decision.

A third type of effect is AFDC’s so-called “entry effect”. Consider panel (a) of Figure 2. A woman whose preferences are represented by an indifference map including the curve through point A optimally would choose not to participate in AFDC. Instead, she would work enough hours so that her labor income exceeds the breakeven level. By contrast, a woman whose indifference map includes the curve through point B optimally would choose to participate in AFDC, working few enough hours so that her labor income is below the breakeven level.

No one with consistent preferences could have an indifference map containing both of these curves. However, panel (b) illustrates the possibility that a woman’s indifference map could contain a single indifference curve that is locally tangent to both points A and B. That means that such a woman would be exactly indifferent between choosing the leisure-consumption bundle at point A, which involves participating in AFDC, and choosing the leisure-consumption bundle at point B, which does not involve welfare participation. A woman whose indifference curve through point B is rotated slightly up and to the right would instead have a

---

12 Of course, state and federal income tax systems might so require, but I am abstracting from those systems for purposes of this discussion. In practice, it is important to account for these systems, as well as the effects of various other means-tested transfer programs (e.g., Food Stamps, the EITC, housing assistance, Medicaid, and so on), but for simplicity I will not do so here. Detailed discussions of these issues appear in Gary Burtless and Jerry A. Hausman (1978), “The Effect of Taxation on Labor Supply: Evaluating the Gary Negative Income Tax Experiment,” *Journal of Political Economy*, Vol. 86, No. 6, Dec., pp. 1103-1130; Michael Keane and Robert Moffitt (1998), “A structural model of multiple welfare program participation and labor supply,” *International Economic Review*, 39: pp. 553-589; and Moffitt (1992 JEL).


14 That is because, for people with consistent preferences, indifference curves never cross.
unique optimum that involved participation in AFDC. One whose indifference curve through point B is rotated slightly down and to the left would strictly prefer not to participate in AFDC. This example illustrates how the budget line non-linearity that AFDC creates will induce some people to reduce income in order to become eligible for welfare assistance. That is the entry effect.

Figure 2: Simplified Budget Sets with No Transfer Program, with an NIT, and with AFDC

Under AFDC, the federal government set the benefit reduction rate $t_{AFDC}$, and states had no discretion to modify it. The benefit reduction rate was 100% from the program’s inception until 1967, at which point it was reduced to 67%. In 1981, Congress restored the benefit reduction rate to 100%.

Under AFDC, states had wide discretion to set the maximum benefit level. Table 1 shows state-specific maximum benefit levels for a family of three in 1996. Even in Alaska, whose $1425 monthly maximum benefit level was greatest among all states, the maximum benefit level is quite small in absolute terms—just 70.8 percent of the official federal poverty line for a family of three. The median monthly maximum benefit level in 1996, $625, was less than 40 percent of the federal poverty line. Mississippi’s value of $200, which was the minimum across states, was less than one-eighth of the poverty line.

---

15 See Moffitt (2003), supra note 11, at 131.
16 See Moffitt (2003), supra note 11, at 131.
17 All dollar figures are expressed in 2010 dollars, using CPI-U to inflate, unless stated otherwise.
Table 1: Maximum Benefit for a Family of Three, 1996

<table>
<thead>
<tr>
<th>State</th>
<th>Nominal 1996 Dollars&lt;sup&gt;a&lt;/sup&gt;</th>
<th>Real 2010 Dollars&lt;sup&gt;b&lt;/sup&gt;</th>
<th>As Share of Poverty Line&lt;sup&gt;c&lt;/sup&gt;</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Minimum 120</td>
<td>200</td>
<td>11.5</td>
</tr>
<tr>
<td></td>
<td>Median 377</td>
<td>625</td>
<td>37.3</td>
</tr>
<tr>
<td></td>
<td>Maximum 923</td>
<td>1,425</td>
<td>70.8</td>
</tr>
<tr>
<td>Alabama</td>
<td>164</td>
<td>270</td>
<td>15.7</td>
</tr>
<tr>
<td>Alaska</td>
<td>923</td>
<td>1,425</td>
<td>70.8</td>
</tr>
<tr>
<td>Arizona</td>
<td>347</td>
<td>581</td>
<td>33.3</td>
</tr>
<tr>
<td>Arkansas</td>
<td>204</td>
<td>343</td>
<td>19.6</td>
</tr>
<tr>
<td>California</td>
<td>596</td>
<td>983</td>
<td>58.2</td>
</tr>
<tr>
<td>Colorado</td>
<td>356</td>
<td>600</td>
<td>40.4</td>
</tr>
<tr>
<td>Connecticut</td>
<td>636</td>
<td>1,030</td>
<td>61.0</td>
</tr>
<tr>
<td>Delaware</td>
<td>338</td>
<td>566</td>
<td>32.4</td>
</tr>
<tr>
<td>DC</td>
<td>415</td>
<td>705</td>
<td>40.3</td>
</tr>
<tr>
<td>Florida</td>
<td>303</td>
<td>506</td>
<td>29.1</td>
</tr>
<tr>
<td>Georgia</td>
<td>280</td>
<td>459</td>
<td>26.8</td>
</tr>
<tr>
<td>Hawaii</td>
<td>712</td>
<td>1,194</td>
<td>59.4</td>
</tr>
<tr>
<td>Idaho</td>
<td>317</td>
<td>531</td>
<td>30.4</td>
</tr>
<tr>
<td>Illinois</td>
<td>377</td>
<td>575</td>
<td>36.1</td>
</tr>
<tr>
<td>Indiana</td>
<td>288</td>
<td>481</td>
<td>27.6</td>
</tr>
<tr>
<td>Iowa</td>
<td>426</td>
<td>688</td>
<td>40.8</td>
</tr>
<tr>
<td>Kansas</td>
<td>429</td>
<td>691</td>
<td>41.1</td>
</tr>
<tr>
<td>Kentucky</td>
<td>262</td>
<td>456</td>
<td>25.1</td>
</tr>
<tr>
<td>Louisiana</td>
<td>190</td>
<td>325</td>
<td>18.2</td>
</tr>
<tr>
<td>Maine</td>
<td>418</td>
<td>731</td>
<td>40.1</td>
</tr>
<tr>
<td>Maryland</td>
<td>373</td>
<td>625</td>
<td>35.8</td>
</tr>
<tr>
<td>Massachusetts</td>
<td>565</td>
<td>905</td>
<td>54.2</td>
</tr>
<tr>
<td>Michigan&lt;sup&gt;*&lt;/sup&gt;</td>
<td>459</td>
<td>782</td>
<td>44.0</td>
</tr>
<tr>
<td>Minnesota</td>
<td>532</td>
<td>863</td>
<td>51.0</td>
</tr>
<tr>
<td>Mississippi</td>
<td>120</td>
<td>200</td>
<td>11.5</td>
</tr>
<tr>
<td>Missouri</td>
<td>292</td>
<td>475</td>
<td>28.0</td>
</tr>
<tr>
<td>Montana</td>
<td>438</td>
<td>732</td>
<td>40.7</td>
</tr>
<tr>
<td>Nebraska</td>
<td>364</td>
<td>605</td>
<td>34.9</td>
</tr>
<tr>
<td>Nevada</td>
<td>348</td>
<td>567</td>
<td>33.4</td>
</tr>
<tr>
<td>NH</td>
<td>550</td>
<td>852</td>
<td>52.7</td>
</tr>
<tr>
<td>NJ</td>
<td>424</td>
<td>678</td>
<td>40.7</td>
</tr>
<tr>
<td>NM</td>
<td>389</td>
<td>652</td>
<td>37.3</td>
</tr>
<tr>
<td>NY**</td>
<td>703</td>
<td>955</td>
<td>55.3</td>
</tr>
<tr>
<td>NC</td>
<td>272</td>
<td>413</td>
<td>26.1</td>
</tr>
<tr>
<td>ND</td>
<td>431</td>
<td>719</td>
<td>41.3</td>
</tr>
<tr>
<td>Ohio</td>
<td>341</td>
<td>585</td>
<td>32.7</td>
</tr>
<tr>
<td>Oklahoma</td>
<td>307</td>
<td>528</td>
<td>29.4</td>
</tr>
<tr>
<td>Oregon</td>
<td>460</td>
<td>785</td>
<td>44.1</td>
</tr>
<tr>
<td>Pennsylvania</td>
<td>421</td>
<td>714</td>
<td>40.4</td>
</tr>
<tr>
<td>RI</td>
<td>554</td>
<td>878</td>
<td>53.1</td>
</tr>
<tr>
<td>SC</td>
<td>200</td>
<td>335</td>
<td>19.2</td>
</tr>
<tr>
<td>SC</td>
<td>430</td>
<td>664</td>
<td>41.2</td>
</tr>
<tr>
<td>Tennessee</td>
<td>185</td>
<td>314</td>
<td>17.7</td>
</tr>
<tr>
<td>Texas</td>
<td>188</td>
<td>314</td>
<td>18.0</td>
</tr>
<tr>
<td>Utah</td>
<td>426</td>
<td>692</td>
<td>40.8</td>
</tr>
<tr>
<td>Vermont</td>
<td>633</td>
<td>988</td>
<td>62.9</td>
</tr>
<tr>
<td>Virginia</td>
<td>354</td>
<td>570</td>
<td>33.9</td>
</tr>
<tr>
<td>Washington</td>
<td>546</td>
<td>892</td>
<td>52.3</td>
</tr>
<tr>
<td>WV</td>
<td>253</td>
<td>434</td>
<td>24.3</td>
</tr>
<tr>
<td>Wisconsin</td>
<td>517</td>
<td>857</td>
<td>49.6</td>
</tr>
<tr>
<td>Wyoming</td>
<td>360</td>
<td>542</td>
<td>34.5</td>
</tr>
</tbody>
</table>

<sup>a</sup> For Wayne County only. **Excluding Suffolk County.

<sup>a</sup> Source: ACF, Aid to Families with Dependent Children: The Baseline, June 1998, Table 5.7

Maximum benefit levels were not always so low in relative terms. Panel (a) of Figure 3 shows the 10th percentile, the median, and the 90th percentile of the distribution of state monthly maximum benefit levels for a family of three for selected dates between 1970 and 1996, in constant 2010 dollars. The 90th percentile in 1996 was $802, far below the 90th percentile for 1970, which was $1,430. Both the median and the 10th percentile also fell substantially over this quarter-century period—from $985 to $541 for the median, and from $557 to $278 for the 10th percentile.

**Figure 3: Trends in Real Maximum Benefit Levels for Family of Three, 1970-1996**

These reductions in monthly maximum benefit levels at these three points in the cross-state distribution were different in absolute-dollar levels, but of course the three points in the distribution start out at different absolute levels. In panel (b) of Figure 3, I scale each of the three time series so that its 1970 level equals 100. The resulting graph shows that trends in the three points in the cross-state distribution tracked each other closely in relative terms. By 1996, each of the three percentiles had lost between 40 and 50 percent of its real 1970 value. Thus, AFDC maximum benefit levels fell throughout the cross-state distribution.

b. Effects on labor supply and welfare participation

As Justice Brandeis surely would have appreciated, the substantial variation of benefit levels across states and over time formed the backbone of the literature on AFDC’s incentive effects. As noted above, basic labor supply theory suggests that a more generous maximum
benefit level should lead to more welfare participation and less paid work.\textsuperscript{18} Moffitt’s (1992) review of the empirical evidence on this question suggests that the literature finds statistically significant effects, but that they are not large. Crucially, he reports that estimated entry effects are sufficiently small so that “most AFDC women would, apparently, be poor even in the absence of the AFDC program.”\textsuperscript{19} Indeed, estimates in Moffitt (1983) imply that characteristics like age and education are more important predictors of labor market and welfare participation outcomes than are welfare program parameters like the benefit reduction rate; he writes that “virtually all of these nonprogram changes have a far greater effect on labor supply than the program changes that have been the focus of the analysis.”\textsuperscript{20}

Given the relatively small labor supply effects of the welfare system, it is natural to wonder what does explain most of the variation over time in AFDC use. One alternative explanation is the mechanical one I discussed in connection with Figure 2. There are many disadvantaged people—in labor economics terms, people with low marginal products—and even in the absence of the welfare system, many of them would work little, or they would earn little despite working considerably more than a little. The expansion of the AFDC system would have made many of these people eligible for public assistance even without affecting their behavior much. If this explanation is largely correct, then the AFDC program induced only small economic costs (over and above the costs of raising the tax revenue needed to fund the program).

A second explanation is that welfare use is sensitive to labor market conditions. There is compelling evidence from some high quality studies that this is true.\textsuperscript{21} By itself, though, this explanation does poorly in explaining long run trends. As Figure 4 shows, the annual average of the number of monthly AFDC recipients rose sharply during the boom of the late 1960s, leveled off during the macroeconomic slowdowns of the 1970s, stayed roughly constant throughout the 1980s, and then began to increase in the late 1980s. Of course, there is no reason why one should seek only a single explanation for a decades-long trend in welfare use, and a combination of the mechanical and labor market explanations seems more promising than either explanation on its own.

\textsuperscript{18} My discussion of welfare’s effects on labor supply here is necessarily brief; for more details, see, e.g., Robert A. Moffitt, “Welfare Programs And Labor Supply,” Chapter 34 in \textit{4 Handbook of Public Economics}, A.J Auerbach and M. Feldstein, eds., 2002.

\textsuperscript{19} Moffitt (1992), supra note 6, at 56.


A third explanation involves what economists would call endogenous preferences and other social scientists might call cultural change. Moffitt (1992) notes, somewhat skeptically, that “It is . . . sometimes argued that there was a reduction in the stigma of welfare receipt over [the late 1960s and early 1970s] as AFDC came to be viewed as a ‘right.’”22 Supreme Court decisions protecting eligibility for AFDC are one potential source, or perhaps only a reflection, of such a reduction in stigma that should be familiar to legal scholars.23

The culture explanation might do some useful work in the more recent period of retrenchment in government’s commitment to assisting welfare families. President Reagan made potent use of the “welfare queen” symbol, and by the 1990s supporting welfare was the sort of thing political consultants might equate with proposing cuts in military spending: some politicians did it, but those with ambitions outside of safely liberal districts did so at their peril. AFDC had been a favorite target among those on the right at least since Charles Murray’s 1984

---

22 Moffitt (1992) at 8.
23 For example, see Goldberg v. Kelly, 397 U.S. 254, 264 (1970) (holding that procedural due process requires an in-person pre-termination evidentiary hearing before welfare benefits may be discontinued, because “termination of aid pending resolution of a controversy over eligibility may deprive an eligible recipient of the very means by which to live while he waits”); Shapiro v. Thompson, 394 U.S. 618 (1969) (holding that states may not deny welfare applications solely on the basis of an applicant’s having resided in-state for less than one year); King v. Smith, 392 U.S. 309, 314 (1968) (finding invalid on statutory grounds an Alabama regulation, which deemed a man a “substitute father” of a mother’s children if they engaged in “‘frequent’ or ‘continuing’ sexual relations”—defined variously as weekly or possibly only semi-annually, and which would have prevented AFDC eligibility by dint of the “substitute father’s” presence in the household unit).
book *Losing Ground*, and the program had become a political pinata by the time of the 1994 Republican Congressional victories.24

Evaluations of President Clinton’s political success often pay homage to his choice to sign PRWORA. The legislation was both the product of a Republican Congress and far outside what would have been considered politically viable, say, ten years before—which itself was a time when Republicans controlled both the White House and the Senate. There is no question that American politics turned sharply against welfare over this period, and so it is not unreasonable to speculate that American culture might have as well. Such an “outward shift in welfare stigma” would provide an additional reason for the large reduction in caseloads, over and above changes in welfare programs’ incentive and entitlement structure and the boom of the 1990s. Unfortunately, the extent of stigma and its impact on behavior is difficult to measure except in the most indirect ways,25 so this hypothesis may be untestable as a practical matter.

c. Categorical eligibility and effects on family structure

Before PRWORA, maximum benefit levels in all states increased when the number of people in the AFDC family unit did. Table 2 shows the incremental benefit received in 1996 when the AFDC unit increased from two to three persons. Measured in 2010 dollars, the incremental maximum benefit level ranged from $33 per month for Mississippi to $204 for Hawaii, with the median being $100.26 The policy of increasing benefits for additional family members came under heavy attack in the years leading up to PRWORA, with critics arguing that even the rather meager incremental benefit levels that appear in Table 2 were enough to induce sizable added fertility. Thus did New Jersey introduced one of the highest-profile state waiver programs: a so-called “family cap” that prevented benefit payments from rising with additions to the family unit.27

---

24 As a mark of just how far against the traditional welfare system politics had swung, then-House Speaker Newt Gingrich proposed removing the children of welfare recipients from their families and placing them in orphanages. Katharine Q. Seelye, “Republicans Plan Ambitious Agenda in Next Congress,” *New York Times*, November 15, 1994. Careful readers might detect a certain tension between this proposal and PRWORA’s focus, see above and below, on the importance of children’s living in stable families with their own parents. Interestingly, as part of his 2012 presidential campaign, Speaker Gingrich has recently suggested that janitorial work in schools would constitute effective job preparation. Maggie Haberman, “Newt: Fire the Janitors, Hire Kids to Clean Schools,” *Politico*, November 18, 2011 8:15PM EST, downloaded from [http://www.politico.com/news/stories/1111/68729.html](http://www.politico.com/news/stories/1111/68729.html) on February 12, 2012. The Speaker has not indicated whether he envisions using orphanages and custodial training as complements or substitutes.

25 See Robert A. Moffitt (1983), “An Economic Model of Welfare Stigma,” *American Economic Review*, Vol. 73, No. 5 (Dec.), pp. 1023-1035, for an attempt. One potential criticism is that “stigma” in Moffitt’s model is essentially any source of model specification error. Thus, for example, the value of Medicaid benefits, on which Moffitt does not have data, contributes to his estimates of stigma. That definition of stigma differs from what I mean to capture here.

26 Incremental maximum benefit levels were generally similar for families having other sizes; for more details, see ACF, *Aid to Families with Dependent Children: The Baseline*, June 1998, Table 5.7, [http://aspe.hhs.gov/hsp/afdc/baseline/5benefits.pdf](http://aspe.hhs.gov/hsp/afdc/baseline/5benefits.pdf).

27 Family cap policies were especially controversial in the 1990s, in part due to some observers’ concerns that they would increase abortions. *See, e.g.*, Jennifer Preston, *Births Fall and Abortions Rise Under New Jersey Family Cap*, The New York Times, November 03, 1998.
Table 2: Real Incremental Maximum Benefit for Increasing Family Size from Two to Three Persons, 1996

<table>
<thead>
<tr>
<th>State</th>
<th>Two Persons&lt;sup&gt;a&lt;/sup&gt;</th>
<th>Three Persons&lt;sup&gt;b&lt;/sup&gt;</th>
<th>Difference</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Minimum</td>
<td>Median</td>
<td>Maximum</td>
</tr>
<tr>
<td></td>
<td>133</td>
<td>434</td>
<td>1,141</td>
</tr>
<tr>
<td>Alabama</td>
<td>190</td>
<td>228</td>
<td>38</td>
</tr>
<tr>
<td>Alaska</td>
<td>1141</td>
<td>1283</td>
<td>142</td>
</tr>
<tr>
<td>Arizona</td>
<td>382</td>
<td>482</td>
<td>100</td>
</tr>
<tr>
<td>Arkansas</td>
<td>225</td>
<td>284</td>
<td>58</td>
</tr>
<tr>
<td>California</td>
<td>666</td>
<td>828</td>
<td>163</td>
</tr>
<tr>
<td>Colorado</td>
<td>389</td>
<td>495</td>
<td>106</td>
</tr>
<tr>
<td>Connecticut</td>
<td>713</td>
<td>884</td>
<td>171</td>
</tr>
<tr>
<td>Delaware</td>
<td>375</td>
<td>470</td>
<td>95</td>
</tr>
<tr>
<td>DC</td>
<td>453</td>
<td>577</td>
<td>124</td>
</tr>
<tr>
<td>Florida</td>
<td>335</td>
<td>421</td>
<td>86</td>
</tr>
<tr>
<td>Georgia</td>
<td>327</td>
<td>389</td>
<td>63</td>
</tr>
<tr>
<td>Hawaii</td>
<td>785</td>
<td>990</td>
<td>204</td>
</tr>
<tr>
<td>Idaho</td>
<td>349</td>
<td>441</td>
<td>92</td>
</tr>
<tr>
<td>Illinois</td>
<td>386</td>
<td>524</td>
<td>138</td>
</tr>
<tr>
<td>Indiana</td>
<td>318</td>
<td>400</td>
<td>82</td>
</tr>
<tr>
<td>Iowa</td>
<td>502</td>
<td>592</td>
<td>90</td>
</tr>
<tr>
<td>Kansas</td>
<td>489</td>
<td>596</td>
<td>107</td>
</tr>
<tr>
<td>Kentucky</td>
<td>313</td>
<td>364</td>
<td>51</td>
</tr>
<tr>
<td>Louisiana</td>
<td>192</td>
<td>264</td>
<td>72</td>
</tr>
<tr>
<td>Maine</td>
<td>434</td>
<td>581</td>
<td>147</td>
</tr>
<tr>
<td>Maryland</td>
<td>406</td>
<td>518</td>
<td>113</td>
</tr>
<tr>
<td>Massachusetts</td>
<td>659</td>
<td>785</td>
<td>126</td>
</tr>
<tr>
<td>Michigan&lt;sup&gt;*&lt;/sup&gt;</td>
<td>516</td>
<td>638</td>
<td>122</td>
</tr>
<tr>
<td>Minnesota</td>
<td>607</td>
<td>739</td>
<td>132</td>
</tr>
<tr>
<td>Mississippi</td>
<td>133</td>
<td>167</td>
<td>33</td>
</tr>
<tr>
<td>Missouri</td>
<td>325</td>
<td>406</td>
<td>81</td>
</tr>
<tr>
<td>Montana</td>
<td>485</td>
<td>609</td>
<td>124</td>
</tr>
<tr>
<td>Nebraska</td>
<td>407</td>
<td>506</td>
<td>99</td>
</tr>
<tr>
<td>Nevada</td>
<td>402</td>
<td>484</td>
<td>82</td>
</tr>
<tr>
<td>NH</td>
<td>668</td>
<td>764</td>
<td>96</td>
</tr>
<tr>
<td>NJ</td>
<td>448</td>
<td>589</td>
<td>142</td>
</tr>
<tr>
<td>NM</td>
<td>431</td>
<td>541</td>
<td>110</td>
</tr>
<tr>
<td>NY&lt;sup&gt;**&lt;/sup&gt;</td>
<td>801</td>
<td>977</td>
<td>177</td>
</tr>
<tr>
<td>NC</td>
<td>328</td>
<td>378</td>
<td>50</td>
</tr>
<tr>
<td>ND</td>
<td>463</td>
<td>599</td>
<td>136</td>
</tr>
<tr>
<td>Ohio</td>
<td>388</td>
<td>474</td>
<td>86</td>
</tr>
<tr>
<td>Oklahoma</td>
<td>331</td>
<td>427</td>
<td>96</td>
</tr>
<tr>
<td>Oregon</td>
<td>549</td>
<td>639</td>
<td>90</td>
</tr>
<tr>
<td>Pennsylvania</td>
<td>459</td>
<td>585</td>
<td>126</td>
</tr>
<tr>
<td>RI</td>
<td>624</td>
<td>770</td>
<td>146</td>
</tr>
<tr>
<td>SC</td>
<td>221</td>
<td>278</td>
<td>57</td>
</tr>
<tr>
<td>SC</td>
<td>528</td>
<td>598</td>
<td>69</td>
</tr>
<tr>
<td>Tennessee</td>
<td>197</td>
<td>257</td>
<td>60</td>
</tr>
<tr>
<td>Texas</td>
<td>227</td>
<td>261</td>
<td>35</td>
</tr>
<tr>
<td>Utah</td>
<td>475</td>
<td>592</td>
<td>117</td>
</tr>
<tr>
<td>Vermont</td>
<td>741</td>
<td>880</td>
<td>139</td>
</tr>
<tr>
<td>Virginia</td>
<td>409</td>
<td>492</td>
<td>83</td>
</tr>
<tr>
<td>Washington</td>
<td>612</td>
<td>759</td>
<td>147</td>
</tr>
<tr>
<td>WV</td>
<td>279</td>
<td>352</td>
<td>72</td>
</tr>
<tr>
<td>Wisconsin</td>
<td>612</td>
<td>719</td>
<td>107</td>
</tr>
<tr>
<td>Wyoming</td>
<td>445</td>
<td>500</td>
<td>56</td>
</tr>
</tbody>
</table>
Categorical eligibility refers primarily to the requirement that a parent either be dead, continuously absent from the home, unable to work due to a disability, or unemployed. Leaving aside unemployment for the moment, these criteria meant that most family units that received AFDC assistance were headed by single mothers. This fact elicited little concern in the program’s early years, because the whole point was to provide support either for orphans or for widows so that single mothers would not have to work.28

Subsequently, though, many critics of the AFDC program pointed to the categorical eligibility requirement as a cause of the growth in female-headed households and out-of-wedlock child bearing, which they saw as an important social ill.29 If AFDC’s categorical eligibility requirement were empirically important, then one would have expected that as real welfare benefit levels fell, marriage rates would rise and divorce rates would fall. Yet trends reported in Blank’s review show that both marriage rates and divorce rates fell steadily after the early 1980s.30 In addition, the birth rate among all unmarried women aged 15-44 increased steadily after the mid-1970s, slowing down around 1990 and peaking around 1994, before most major welfare reforms took hold.31 Thus the theory advanced by AFDC’s critics fits the time series trends rather poorly.32

Simple time series evidence can be only the jumping-off point for a phenomenon as complex as fertility and marital behavior, of course. But the micro-econometric literature also shows little reason to believe that welfare programs caused much of the changes in family structure with which AFDC’s opponents charged the program. Moffitt (1992) has written that “the econometric estimates of family structure effects are not large enough to explain long-run declines in marriage rates and, in any case, are incapable of explaining recent upward trends in female headship because welfare benefits have been declining.”33 Further, “the welfare system does not appear to be capable of explaining most of the long-term trend or any of the recent trend of increasing numbers of female-headed families in the United States.”34 And work done after Moffitt’s 1992 review showed, as Hilary Williamson Hoyes has put it, that estimates aren’t even statistically significant once one controls for state and/or individual fixed effects (so that the

29 Id.
30 Rebecca Blank (2002), supra note 6, at 1154, Figure 5.
31 Blank notes that the early-1990s slowdown “in the nonmarital birth rate is evident among both black and white women, and among teens and older women,” Id.
32 Whatever the causal impact of the requirement of parental deprivation, one result of this debate was the creation of the AFDC-Unemployed Parent program in 1990. I do not discuss this aspect of AFDC in detail here, because it involved a relatively small share of the caseload. For details, see, e.g., Hoyes (1996).
33 Moffitt (1992), supra note 6, at 56-57.
34 Moffitt (1992), supra note 6, at 57.
benefit level variation identifying the estimates is not the result of permanent differences across states or individuals).

One paper that does appear to find substantial welfare-induced fertility effects is by Mark Rosenzweig. Rosenzweig uses a different measure of welfare generosity than most other authors in this literature. Most authors had used the state maximum monthly benefit variable for the year in which a woman was observed, Rosenzweig uses the average maximum benefit level, for a woman with two children, over the nine year period when a woman was aged 12-20. Rosenzweig’s rationale for using this measure of benefit generosity, rather than the contemporaneous year-specific maximum benefit level, is that the average value better captures the benefit available over the long run, which should be more relevant than year-to-year values for a woman evaluating the level of support she can expect from public assistance over the extended period it takes to raise a child. His data, from the National Longitudinal Survey of Youth, include information on overlapping birth cohorts—those from 1958-1965—so Rosenzweig is able to control for state fixed effects as well.

Rosenzweig uses a multinomial logit model that groups women’s combined marriage and fertility outcomes, over the period when they are aged 14-22, into categories involving (i) no births, (ii) only marital births, or (iii) any premarital births, defined as a birth that occurred more than six months before the woman got married, if she did. His key regressor is the average maximum benefit level variable discussed above. He then uses his multinomial logit coefficient estimates to predict the change in the share of sample members who would wind up in each of categories (i), (ii), and (iii) had the AFDC maximum benefit level been $130 lower (measured in 1985 dollars). Leaving aside some technical details that can be glossed over for purposes of my discussion, Rosenzweig finds that this reduction in benefit generosity would reduce the share of his overall sample that ever has a premarital birth from 0.104 to 0.073—a reduction of 30%.

36 Given that Rosenzweig includes fixed effects, his identifying variation comes from state-level changes in average benefit generosity. Taking changes in averages will tend to focus estimator attention on changes in persistent benefit levels, rather than changes in year-to-year benefit levels. For further discussion of issues related to long run versus year-to-year changes in a regressor, see Terra G. McKinnish, Model Sensitivity in Panel Data Analysis: Some Caveats About the Interpretation of Fixed Effects and Differences Estimators, unpublished and undated paper, available at http://stripe.colorado.edu/~mckinnis/fe053100.pdf.
37 Rosenzweig’s sample contains women born at different times in any given state, and real state maximum benefit levels vary over time. Cohort-based differences in state-specific average maximum benefit levels provide the variation in benefits necessary to both (i) use the average maximum benefit level a woman faces as a regressor, and (ii) include state fixed effects. He also includes birth cohort dummies.
38 Rosenzweig, at S21, Table 5.
Among those in Rosenzweig’s low-parental income subsample, the corresponding change is from a baseline of 0.170 to 0.096, which is a drop of 44%.

Rosenzweig’s estimates suggest considerably larger effects than do others in the literature. Perhaps this is because they are identified from lifecycle rather than noisy year-to-year variation, as Rosenzweig argues. But mis-specification is a plausible alternative explanation. As I noted above, (i) Rosenzweig’s measure of long run welfare benefits is the (unweighted) average state maximum benefit level during the period when the sample member is aged 12-20, and (ii) his measure of pre-marital fertility is a dummy equal to one for sample members who have any pre-marital births in the period when they are 14-22. I see three potential problems with this approach.

The first problem has to do with what one means by a long run benefit measure. Even assuming that women know the contemporaneous benefit level in their states, as virtually all authors implicitly do, it is not obvious that sample members know the benefit levels their state would have years later. In other words, does it make sense to assume that they have perfect foresight? If sample members just assume—however wrongly—that the current year’s benefit level will persist forever, then the contemporaneous benefit level is the right one in any given year, and Rosenzweig’s model is mis-specified. Thus, Rosenzweig’s econometric model, and thus his empirical results, seem to hinge importantly on an unstated but hyper-informed expectational model. One alternative would be to model expectations of future benefit levels using an adaptive expectations framework. In such a framework, expectations of future benefit levels would be a function of the sequence of past benefit levels, with declining weight placed on those benefit levels further in the past. Yet Rosenzweig includes only two lags from the perspective of a 14-year-old sample member, since no maximum benefit levels are included before the sample member turns 12. Moreover, in forming his average, he puts the same weight on all years’ benefit levels.

The second problem has to do with which years’ maximum benefit levels are missing from Rosenzweig’s measure of maximum benefit levels. Rosenzweig excludes maximum benefit levels from the years when a sample member is aged 21 and 22, yet fertility behavior in those years is part of his dependent variable. Thus the way he defines his welfare benefit measure has the very odd property that it involves assuming women (i) have perfect foresight concerning future benefits at the same time that (ii) in two of the nine years he studies, they completely ignore contemporaneous benefit levels.

The third issue is Rosenzweig’s choice to aggregate fertility behavior over a lengthy, nine-year period. A more reasonable approach would have been to estimate a discrete-time hazard model. In such a model, each woman has a separate observation each year that she is at risk to have a pre-marital birth (and those who have already had such a birth exit the sample once they do). This approach would certainly be no less econometrically correct than Rosenzweig’s aggregate approach. With regard to my first specification criticism above, a discrete-time

---

41 Rosenzweig, at S21, Table 5.
42 Indeed, that statement is a bit charitable. The nine-year aggregate behavioral outcome is the result of nine individual years of behavior. It is not clear what year-by-year model would result in Rosenzweig’s nine-year-aggregate model. Of course, every model has its unrealism. But it is not clear why Rosenzweig could not have estimated a year-by-year hazard model; for an example, see Jeff Grogger and Stephen G. Bronars, The Effect of
hazard model would have forced Rosenzweig to consider explicitly what model to use for sample members’ expectations concerning long run maximum benefit levels. With regard to my second criticism, this approach would have made it obvious that—whatever model of expectations Rosenzweig used—excluding the contemporaneous benefit level cannot be justified.

In sum, it is difficult to know what to make of Rosenzweig’s empirical results. I certainly agree with his basic point—that long run variation in welfare benefits should matter more in determining fertility decisions than should year-to-year variation. But the problems with his econometric implementation raise the possibility that his results hinge importantly on measurement and specification choices that both (i) are difficult to justify and (ii) plausibly drive his results. I do not think Rosenzweig’s paper importantly changes the basic conclusion from the pre-existing literature on welfare and family structure—that there were likely relatively small effects in general. And note again that time series patterns show increases in non-marital fertility (see Error! Reference source not found.) at the same time as real welfare benefit levels were falling (see Figure 3). From the perspective of understanding trends in non-marital fertility, then, sizable welfare program effects on fertility and family structure would simply increase the scope for non-welfare sources of change.

d. Other features

I have glossed over many details in how the AFDC program worked from the perspective of beneficiaries; I will mention some of these here in the interests of completeness. A number of other safety net programs interacted with AFDC. The two most notable are Food Stamps and Medicaid. AFDC participation often brought automatic eligibility for the Food Stamps program, which also operated like an NIT program with no possibility of paying positive tax. Let $G_{fs}$ be the Food Stamps maximum grant and let $t_{fs}$ be the Food Stamps benefit reduction rate. The Food Stamps program “taxed” AFDC benefits in the sense that it treated AFDC payments as income for Food Stamps benefit-computation purposes. Thus, an AFDC family’s total payment from the two programs, assuming it was positive, equaled

$$P_{\text{total}} = G_{fs} + (1 - t_{fs})[G_{AFDC} + (1 - t_{AFDC})Y],$$

and setting $t_{AFDC} = 1$, this turns into

$$P_{\text{total}}^* = G_{fs} + (1 - t_{fs})G_{AFDC}.$$ 

Thus, given that the Food Stamps benefit reduction rate, $t_{fs}$, equaled 30 percent, the total benefit package equaled the maximum Food Stamps benefit plus 70 percent of the maximum AFDC benefit payment. Because Food Stamps benefits were substantial by comparison to maximum AFDC benefits, even for relatively generous states, the Food Stamps part of the benefit was an important part of the traditional welfare system’s transfer payment. For example, the maximum monthly food stamps allotment for a family of three in 1996 was $435 when

---


43 This program has been renamed and is now called the Supplemental Nutrition Assistance Program, or SNAP.

44 See Committee on Ways and Means (1994), Green Book, chapter 10 (available at http://aspe.hhs.gov/94gb/sec10.txt). Takeup was very high, with 87.3% of AFDC units also receiving Food Stamps payments in 1992; see Table 10-27.
measured in 2010 dollars.\textsuperscript{45} The median real AFDC benefit for a family of three in 1996 was New Mexico’s $541. Thus, a family of three in New Mexico could receive as much as $813 per month in 1996.\textsuperscript{46}

Because the food stamps benefit is uniform in the (contiguous) United States, food stamps benefits greatly increased the net value of participating in AFDC in states with very low AFDC benefits. For example, the lowest maximum monthly AFDC benefit level for a family of three in 1996 was Mississippi’s $167. A Mississippi family receiving this benefit would receive an AFDC benefit payment of $167 plus a food stamps payment of $385, for a total of $552. For this family, the food stamps benefit accounts for approximately 70 percent of the total benefit payment from the two programs. On the other end of the spectrum, consider a family in Vermont, whose 1996 maximum monthly AFDC benefit of $912 was the highest value in the (contiguous) United States. A family that received this payment would have received a food stamps benefit of only $161, which would have constituted only about 15 percent of the $1,073 total benefit payment from the two programs. These calculations show that the relative contribution of the food stamps program to AFDC families’ resources varied dramatically according to the states in which the families lived.

Hoynes and Schanzenbach (2012) have recently studied the labor supply effects of introducing the Food Stamps program.\textsuperscript{47} While some of their estimates are imprecise, they find very large labor supply responses. For example, their Table 2 (on page 157) shows that among single parent households, the average annual number of hours worked by the household head falls by 183. Scaling this effect to account for the fact that not all women in this group participated in food stamps, their results imply a reduction of 505 hours of work per year, nearly half the total of 1,068 annual hours worked among heads of single parent households in Hoynes and Schanzenbach’s estimation sample. On the other hand, average earnings by the head of these households fell by only $533 (in 2005 dollars), or $1,472 after scaling to account for the food stamps participation rate.

But even the scaled drop represents only a 10% reduction relative to mean earnings in the sample, which were $14,194. A 50% drop in hours can be associated with a 10% drop in earnings only if the group of people reducing hours worked has especially low wages. To verify this observation, I divide the $1,472 drop in earnings implied for food stamps participants by the corresponding drop in labor supply of 505 hours. The result is $2.91 per hour, suggesting that while the food stamps program may have entailed large disincentives in hourly terms, these effects must have been concentrated among precisely those workers for whom the distortion

\textsuperscript{45} According to Table 16.6 of the 1996 Green Book, the maximum monthly allotment was $313 for a family of three persons living in the contiguous United States (allotments were higher in Alaska, Hawaii, Guam, and the Virgin Islands). Multiplying $313 by the 2010 CPI-U index number of 218.056 and dividing by the 1996 CPI-U index number of 156.9 yields $435.

\textsuperscript{46} This is the result of adding the food stamps maximum allotment of $435 to the maximum AFDC maximum benefit of $541 and then subtracting 30 percent of the latter benefit.

\textsuperscript{47} The food stamps program was introduced over time on a county-by-county basis between 1963 and 1975. To estimate the effects of introducing the program in new counties, Hoynes and Schanzenbach use labor market outcomes data covering the period 1967-1978 from the Panel Study on Income Dynamics. Their Figure 1, on p. 153, shows that roughly 40% of counties already had food stamps programs in 1967, with 100% of counties having them by 1974.
would have been especially inexpensive in the sense of reducing income from the labor market.\textsuperscript{48,49}

AFDC participation also brought automatic eligibility for Medicaid for both parents and children in the AFDC family unit. Given the rising cost of health care, this was another important part of the welfare system, especially for people with predictably high medical costs for either adults or children in the AFDC family unit. Thus, the threat of losing Medicaid eligibility served as a deterrent to leaving welfare. Beginning in the 1980s, Congress changed the Medicaid rules to allow states to provide eligibility to children even in the absence of AFDC participation.\textsuperscript{50} These “Medicaid expansions,” which have continued to the present time, thus reduced the work disincentive caused by Medicaid eligibility for families receiving welfare.\textsuperscript{51}

The AFDC program also allowed certain expenses to be “disregarded” from income for benefit-computation purposes. So-called earnings disregards existed for work expenses like transportation as well as for child care costs.\textsuperscript{52} In addition, the Family Support Act of 1988\textsuperscript{53} included both work and training requirements.\textsuperscript{54}

2. The Federal-State Relationship in AFDC

AFDC and its predecessor ADC were always joint federal-state programs. Under ADC, the federal government would reimburse states for one-third of benefit payments made, up to $18 per month for the first child in a home and up to $12 per month for each additional child.\textsuperscript{55} Thus,

\begin{itemize}
\item \textsuperscript{48} In 1973, the monthly maximum food stamps benefit level for a family of four was $114 (see data file posted by Robert Moffitt at http://www.econ2.jhu.edu/people/moffitt/datasets.html and documented at http://www.econ2.jhu.edu/people/moffitt/ben_doc.pdf). Converting to 2005 dollars for comparability with Hoynes and Shanzenbach’s figures, this amounts to $501 (I used the CPI calculator at http://www.bls.gov/data/inflation_calculator.htm to do the conversion). This is considerably greater than the monthly earnings reduction corresponding to a drop in annual earnings of $1,472. Another way to put all this is to observe that the hourly wage corresponding to mean earnings of $14,194 and mean hours of 1,068 is $13.29—far in excess of the implied average wage for women whose labor supply was importantly affected by the introduction of the food stamps program.
\item \textsuperscript{51} Other programs interacting with AFDC included child care subsidies and both state and federal EITC programs.
\item \textsuperscript{52} Some authors have treated these disregards as if they reduce the effective benefit reduction rate. However, this is appropriate to do only if the expenses in question are per se valued by welfare recipients. This seems plausible for high quality child care, but not obviously for, say, bus fare.
\item \textsuperscript{53} Public Law 100–485.
\item \textsuperscript{54} See 1996 Green Book, at 388.
\item \textsuperscript{55} According to data posted by the St. Louis Federal Reserve Bank at http://research.stlouisfed.org/fred2/data/CPIAUCNS.txt, the current price index for urban consumers was 13.7 in
the net cost to states of the first dollar of assistance was 67 cents but rose to a full dollar when benefit payments exceeded the match cap.

Section 411 of the Social Security Act of 1965, which created the Medicaid program, radically altered this system in two ways. First, the 1965 Act allowed states to use matching rates from the newly created Medicaid program (formally, the “Federal medical assistance percentage”), rather than one-third, as the federal matching rate for reimbursement of state AFDC expenditures. This was an important difference because the 1965 Act’s language ensured that the Medicaid matching rate would never be less than 50 percent and could be as high as 83 percent. As of 1996, the match rates varied between 50 and 78 percent with the match-rate formula set so that states with lower per capita incomes received greater matching rates. Second, the 1965 Act eliminated the cap on the amount of benefit payments that the federal government would match.

The net effect on the federal-state relationship of these 1965 changes, then, was to convert the AFDC program from one with a capped match at a rate of one-third to an uncapped match at a rate of at least one-half, and possibly as high as five-sixths. The Food Stamp program came into being at roughly the same time. To see why these changes matter, consider the first two columns of Table 3, which report the cost to a state of providing a dollar of assistance in the pre-1965 period and the period between 1965 and PRWORA’s enactment in 1996.

Table 3: State Cost of Providing Another Dollar of Assistance to Recipient Family

<table>
<thead>
<tr>
<th>Maximum Benefit Level</th>
<th>Pre-1965</th>
<th>1965 – 1996</th>
<th>1996 and later</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>State with Match Rate of 50 Percent</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Less than federal match cap</td>
<td>0.67</td>
<td>0.71</td>
<td>1.43</td>
</tr>
<tr>
<td>Greater than federal match cap</td>
<td>1.00</td>
<td>0.71</td>
<td>1.43</td>
</tr>
<tr>
<td><strong>State with Match Rate of 80 Percent</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Less than federal match cap</td>
<td>0.67</td>
<td>0.29</td>
<td>1.43</td>
</tr>
<tr>
<td>Greater than federal match cap</td>
<td>1.00</td>
<td>0.29</td>
<td>1.43</td>
</tr>
</tbody>
</table>

The table’s first row considers a state whose pre-1965 maximum benefit level was less than the federal cap for matching purposes. To increase an AFDC family’s income by one dollar cost the state 67 cents, since the federal government would contribute one-third of the one dollar of benefits paid. The second row considers a state whose maximum benefit level exceeded the federal cap for matching funds. This state would bear the full cost of providing another dollar of assistance.

July 1935 and 225.922 in July 2011. Dividing the 2011 figure by the 1935 figure yields a ratio of 16.49. Multiplying this by $18 and $12 yields $297 and $198, which are non-trivial figures relative to contemporary benefit levels.

57 To take advantage of this option, states had to start a Medicaid program. Id.
58 Id., § 1905(b)(1), at 352.
60 Id.
Now consider the situation between 1965 and 1996, for states using the Medicaid match rate. Since such states faced no cap, figures in the top two rows of Table 3 are the same, which concern states whose match rates were 50 percent, as was the case for 11 states plus the District of Columbia in 1996. Recall that the federal Food Stamp program, which was born in 1965, implicitly taxes welfare benefits at 30 percent. Thus, after 1965, an AFDC family would have to receive $1/0.7 = $1.43 in welfare payments to increase its resources by one dollar. States with a 50 percent Medicaid match rate had to bear only half of this cost, or about 71 cents.

For states whose pre-1965 maximum benefit level was less than the cap for federal matching, the combined effect of introducing the Food Stamp program and switching to the Medicaid match rate was minor—an increase in the price of providing assistance from 0.67 to 0.71. But for states with maximum benefit levels above the cap for federal matching, the combined effect was a substantial reduction in the price of providing assistance, from 1.00 to 0.71. Moreover, by eliminating the cap altogether, the 1965 Act gave states with maximum benefit levels below the pre-existing cap an incentive to increase their maximum benefit levels. For states with Medicaid match rates above 50 percent, these effects could be magnified considerably. The bottom two rows of Table 3 show that a state with a Medicaid match rate of 80 percent would have had a price per dollar of assistance equal to just 29 cents between 1965 and 1996.

It is interesting to consider the effects these changes might have had on state benefit levels. The obvious prediction is that relatively more generous states, and states with greater federal match rates, would have increased their maximum benefit levels considerably. To my knowledge, there is no systematic empirical evidence as to whether these predictions were borne out. But the question is intriguing, because, as Figure 5 shows, AFDC caseloads began to take off not long after the 1965 changes. (I discuss other features of Figure 5 in further detail below.)

Figure 5: Annual Average of Monthly Number of AFDC Recipients (in 1000s)

![Graph of Figure 5](image)


I turn now to one of the least notorious aspects of PRWORA: its radical alteration of states’ marginal cost of providing assistance to welfare recipients. PRWORA did not just allow states to implement the changes in the relationship between individual behavior and eligibility/benefit levels discussed above. It also converted the cash welfare system from a federal-state partnership with matching at the margin into a simple block grant. 63 States are not only allowed to design their programs largely as they please—they are also on the hook to pay for the marginal costs of program generosity. The final column of Table 3 illustrates the role of this change in welfare program structure. Since the federal Food Stamp program continues to exist for TANF recipients, 64 its implicit taxation of cash assistance payments continues as well. Thus, a one-dollar increase in cash assistance delivers only 70 cents of additional disposable income to the recipient family. To deliver a full dollar at the margin thus requires \( \frac{1}{0.7} = \$1.43 \) in additional state spending. Table 3’s final column thus shows that PRWORA’s block-granting of federal TANF funds doubled the price of providing cash assistance for a state with a 50 percent match rate, while quintupling it for a state with an 80 percent match rate.

All of this is especially notable because states are allowed to spend large amounts of their TANF block grant funds on things other than direct benefit payments. As a consequence, the TANF program is, roughly speaking, a system in which the federal government hands lump sums to the states and tells them to do as much or as little cash assistance as they please. 65 Under TANF, states are allowed to spend block-grant funds on a variety of non-transfer purposes, including child care, transportation, and job training. 66 Between these structural changes and time limits, it should come as no surprise that cash assistance payments have plummeted since PRWORA. Figure 6 shows the steep decline in real cash payments under TANF that began following PRWORA. By 2009, less than a third of all payments made under the TANF program were made in the form of cash transfers to recipients. 67

---

63 See §103 of PRWORA, 110 Stat. 2113.
64 Among its non-AFDC-related aspects, PRWORA greatly restricted Food Stamp eligibility for able-bodied adults without children; see 7 U.S.C. 2015(o)(2) , added by §824(a) of PRWORA, 110 STAT. 2323, which places a three-month time limit on SNAP receipt, during any 36-month period, for adults aged 50 or younger who do not satisfy state work requirements. For more discussion, see, e.g., Betsey A. Kuhn, Michael LeBlanc, Craig Gundersen (1997), “The Food Stamp Program, Welfare Reform, and the Aggregate Economy,” American Journal of Agricultural Economics, Vol. 79, No. 5, Proceedings Issue (Dec.), pp. 1595-1599. These restrictions were temporarily lifted by §101(e) of the American Recovery and Reinvestment Act of 2009; see 123 STAT. 121.
65 Section 103 of PRWORA, 110 Stat. 2113, codified at 42 U.S.C. 609, does penalize states for failing to “maintain certain level[s] of historic effort.” In addition, states can always use their own funds to pay benefits to time-limited people, though they bear the full marginal cost of providing such benefits.
66 See section II.C.2 for more on diversion.
67 Given that the monthly caseload fell roughly 60 percent over the period between 1996-2009, this change is not too surprising. It is an interesting fact about the block-granting of federal payments under PRWORA that, while inflation worked out to about 39 percent over this period, the federal government transfers the same number of nominal dollars to states even as caseloads have fallen 60 percent.
There is a sizable literature concerning the determinants of state maximum benefit levels. An excellent paper in this literature is David Ribar and Mark Wilhelm’s.\textsuperscript{68} Their main objective is to estimate state-level price and income elasticities of providing cash assistance. Here, the price of providing cash assistance is defined as the product of (i) the state share of each dollar of assistance provided\textsuperscript{69} and (ii) the share of the state population that receives AFDC. Income is defined as per capital total annual personal income.\textsuperscript{70} Ribar and Wilhelm review earlier work, which they characterize as producing “an unacceptably wide range of estimates.”\textsuperscript{71} They argue that this wide range is the result of mis-specification, and they present evidence that specification


\textsuperscript{69} This share is one minus the federal match rate discussed above; see Ribar and Wilhelm (1999) at 97.

\textsuperscript{70} See Ribar and Wilhelm (1999) at 101, Table 2.

\textsuperscript{71} Ribar and Wilhelm (1999), at 96.
problems are solved when one includes state fixed effects, year fixed effects, and state-specific linear trends. Their own results suggest price elasticities, of -0.14 to 0.02, and income elasticities that are generally small, though with confidence bounds “extending from 0.11 to 0.82.”

They conclude that “welfare generosity is much less sensitive to economic changes than many analysts had previously supposed.”

A number of papers have shifted attention away from estimating conventional demand functions. For example, Moffitt (1999) studies the factors that led to the state and federal welfare policy contractions discussed in this chapter, writing from an explicitly public choice-oriented point of view that focuses on interactions between voter preferences, the labor market, and demographic change. He concludes that plausible arguments can be made that welfare reform in the U.S. in the 1990s was strongly influenced by increases in public expenditures on welfare in the late 1980s and early 1990s, just before the major contractionary reforms; by reductions in real incomes and wages of the poor, and their employment rates as well, over the 1980s; and changes in the characteristics of welfare recipients and the poor, particularly the rise in unmarried single motherhood.

Another collection of papers has focused on estimating the spillover effects of states’ welfare policies onto other states, an issue of obvious interest from a federalism perspective. Examples include Figlio, Kolpin and Reid (1999) and Saavedra (2000). Figlio, Kolpin and Reid (1999) report estimates that suggest states’ benefit levels move in tandem with their neighbors’ benefit levels, with own benefits rising and falling roughly dollar-for-dollar as neighbors’ rise and fall. In addition, Figlio et al. present results (in their Table 3, at 450) suggesting that states pay more attention to reductions in their neighbors’ benefits than to increases, so that “states are more concerned about being ‘left ahead’ in welfare benefit levels than they are about being ‘left behind.’” Saavedra (2000) estimates a model constructed by Wildasin (2000) and reports results that “suggest states behave strategically when they set AFDC benefits. The estimated slope parameter of the reaction function is positive and significantly different from zero, indicating that AFDC benefits in any given state are positively affected by the AFDC benefits in neighboring states.”
This topic of state fiscal competition is intimately bound up with the issue of welfare migration, which itself is an especially important topic in the federalism context. If migration is relatively inexpensive, then states that provide generous welfare benefits will find themselves attracting would-be beneficiaries. Such welfare magnetism would raise two important concerns. First, it suggests the possibility of a “race to the bottom” in benefit setting. Second, since many estimates of welfare policy’s behavioral effects on non-migration outcomes are identified primarily from variation in state-level welfare policy variables, endogenous location raises the possibility that much of this literature is subject to a serious endogeneity problem. Simply put, if state benefit levels are an important determinant of where people choose to live, then these benefit levels are choice variables that cannot be treated as exogenous. In this section, I briefly discuss both theoretical and empirical issues related to this so-called “welfare magnet” problem.

In the leading theoretical treatment of welfare migration, Charles Brown and Wallace Oates construct a two-jurisdiction general equilibrium model that has both poor and non-poor agents who can choose the jurisdiction in which to live. They show that the welfare magnet problem can be conceived as a garden variety externality, because each jurisdiction in a decentralized welfare system can offload some of its poor residents onto the other jurisdiction by reducing its welfare benefit level. A “race to the bottom” might occur, resulting in a lower equilibrium benefit level in each jurisdiction than there would be if (i) no one could migrate, or (ii) the benefit level were set centrally.

There are several potential system-design responses to this negative externality. One is to simply run a centralized system—as with, say, Medicare or Social Security. A second approach is to use the kind of hybrid system the U.S. had under AFDC’s federal match of state benefit payments, with federal subsidies counteracting the race to the bottom incentives. Third, one might restrict migration’s usefulness to the migrant in securing cash assistance. Brown and Oates note that the English Poor Laws took this approach: under the Law of Settlement and Removal of 1662, “church wardens and overseers were directed to remove to his ‘home’ parish any newcomer likely to become a burden to his adopted parish unless the new arrival could give surety that he would not become indigent or rented property of the value of ten pounds per year or more.”

Interestingly, attempts to restrict the welfare eligibility of people who move across state lines twice have been rebuffed by the U.S. Supreme Court. In Shapiro v. Thompson, the Court struck down Connecticut, Pennsylvania and District of Columbia statutes that imposed a one-year waiting period before in-migrants could receive welfare benefits. Justice Brennan’s opinion for the Shapiro Court discusses at length the state defendants’ concerns related to the long run burdens of serving as welfare magnets. Nonetheless, the Court held that migration across state

81 As a corollary, such attractor states would have to levy higher taxes than repellor states, which will drive away better off citizens, which would cause the tax base to shrink and exacerbate the problem.
85 Id., at 627-629.
lines is a constitutional right\textsuperscript{86} and observed that “we do not perceive why a mother who is seeking to make a new life for herself and her children should be regarded as less deserving because she considers, among others factors, the level of a State's public assistance.”\textsuperscript{87}

In \textit{Saenz v. Roe},\textsuperscript{88} the Court confronted a California statute that limited the welfare payment an in-migrant could receive to the payment in her previous state of residence, when that payment would be less than California’s, during the in-migrant’s first year of California residence. The Court held that such a limitation was impermissible in light of the 14\textsuperscript{th} Amendment’s Privileges Or Immunities clause\textsuperscript{89} and that PRWORA’s explicit authorization of such a policy was unavailing, since “Congress may not authorize the States to violate the Fourteenth Amendment.”\textsuperscript{90}

Of course, a race to the bottom needs state runners. The Figlio \textit{et al.} (1999), Saavedra (2000), and Baicker (2005) papers discussed above do suggest that states pay attention to their neighbors’ benefit levels. Moreover, Baicker (2005, at 540) reports empirical evidence consistent with the view that “the states with the greatest potential in-migration are those that are most concerned with staying in line with their neighbors’ spending.”

A different angle from which to approach the race to the bottom question is to ask whether there is so much welfare-induced migration that informed and rational state policy makers would adjust benefit levels substantially. The literature on the extent of welfare-induced migration spans many decades and can usefully be divided into roughly three waves. Moffitt 1992 characterizes the first wave, which used data from the late 1960s to early 1970s as finding “rather weak or inconsistent effects of benefits on migration.”\textsuperscript{91} But as he notes, this literature was “severely hampered by a high level of data aggregation and by a consequent inability to disaggregate by individual characteristics, often even by female headship and AFDC receipt.”\textsuperscript{92}

The second wave of welfare migration studies\textsuperscript{93} used micro-data and typically found “positive and significant effects of welfare on residential location and geographical mobility.” At 34. But Moffitt 1992 notes that these studies face their own methodological challenges. For example, they tend to take initial location as given, even though this location itself is the result of

\textsuperscript{86} \textit{Id.}, at 631 (“freedom to travel throughout the United States has long been recognized as a basic right under the Constitution”) (citing \textit{United States v. Guest}, 383 U.S. 745, 757-758 (1966) (quotation mark removed).

\textsuperscript{87} \textit{Shapiro v. Thompson}, 394 U.S. 618, 632.


\textsuperscript{89} \textit{Id.}, at 503 (citing also, at 502, to both the majority and dissenting opinions in \textit{the Slaughter-House Cases}, 16 Wall. 36 (1872), for the proposition that “the right of the newly arrived citizen to the same privileges and immunities enjoyed by other citizens of the same State . . . is protected not only by the new arrival's status as a state citizen, but also by her status as a citizen of the United States”).

\textsuperscript{90} \textit{Id.}, at 507.

\textsuperscript{91} Moffitt (2002), \textit{supra} note 6, at 34.

\textsuperscript{92} Moffitt (2002), \textit{supra} note 6, at 34.

choice. Bruce Meyer has further noted that conditioning on welfare receipt or poverty status when choosing study samples may introduce another important source of bias.94

The final wave of welfare migration studies consists of several that have tried to use either special geographic or lifecycle features to identify effects. Terra McKinnish has written two important papers on welfare migration. Both of these papers take advantage of the facts that (i) welfare benefit levels are set at the state level, while (ii) it will be cheaper to move from a low-benefit to a bordering high-benefit state when a woman lives near the border than when she lives far from it. Using county-level AFDC expenditures as her dependent variable, McKinnish (2005) finds that a $100 reduction in a bordering state’s welfare benefit level is associated with a 4.9% drop in AFDC expenditures in a state’s own county when the own county lies within 25 miles of the bordering state, relative to counties that are further from the bordering state.95 In McKinnish (2007), she uses micro-data to re-do this analysis; the resulting estimates are consistent with the existence of welfare migration, though they are not generally statistically significant.96

In my own paper, Migration, the Life Cycle, and State Benefits: How Low Is the Bottom?,97 I pointed out that the incentive to move to a higher-benefit state is much greater for women with younger children, since they can expect to have more years of future welfare eligibility. For example, under AFDC rules, a woman whose youngest child is newborn would have 18 years of remaining eligibility, whereas a woman whose youngest child is 17 would have just one year. The incentive to move for greater benefit levels is much stronger in the former case.

Moreover, I also pointed out that previous studies on welfare migration had tended not to use valuable information from the welfare dynamics literature. This literature shows that AFDC could be thought of as two very different programs stapled together. The first program was a temporary assistance program for women who had recently experienced a financially adverse event—typically a divorce, lost job, or new birth outside of marriage. These women would participate for relatively short periods of time until they got back on their feet. The second program embedded in AFDC was a long-term income support program, which provided assistance to single mothers as long as they had minor children.

Once one views the issue of welfare migration from a dynamic perspective, it becomes clear that migration incentives are likely to matter only for those women affected by the second, more permanent type of AFDC support. Because data sets do not include an explicit “how-long-would-you-spend-on-welfare” variable, researchers have no choice but to find proxies for such a variable, and many used high school dropout status as an indicator of attachment to the welfare

system. But dropout status is a coarse filter for these purposes, as Mary Jo Bane and David Ellwood’s 1994 book Welfare Realities: From Rhetoric to Reform, which shows considerable heterogeneity according to whether high school dropouts have ever been married.

Bane and Ellwood report simulation results suggesting that, among those who ever have a spell on welfare, never-married high school graduates and ever-married high school dropouts would receive cash welfare for roughly 6 years over a 25-year window. By contrast, never-married high school dropouts would spend an average of at least 10 years on welfare, while an ever-married high school graduate would spend just four years on welfare over this period. And these figures mute the long run differences, since both dropout and never-married status are positively associated with the event of commencing a welfare-use spell in the first place. Thus, there is considerable value in focusing on never-married dropouts when trying to measure the effects of welfare benefit levels on migration choices, given the very important role of expected future program use in this decision.

I reported results from 1980 and 1990 Census microdata samples. For the 1980 sample, I found that single mothers’ out-migration from a state between 1975 and 1980 is quite sensitive to states’ own benefit levels, with the sensitivity greatest among those who are most likely to make extensive use of the welfare system. These results are consistent with the existence of lifecycle welfare migration, though results for the 1990 sample are too unclear to draw any conclusions.

An additional contribution of my paper was to construct a metric that allows one to assess the magnitude of any welfare migration effects. Previous work had tended to focus on measures like the share of migration that was caused by cross-state welfare benefit differentials. Such measures do not map into answering the question of how much an impact welfare migration could have on (either optimal or actual) state benefit levels. To answer this question, I constructed a simple model in which each state chooses its benefit level to maximize an objective function that depends positively on transfers to the poor, positively on non-welfare government expenditures, and negatively on taxes. I then show how to compare the state’s optimal welfare benefit level in the world where state-to-state welfare migration is allowed to the optimal welfare benefit level in the world where it is not (i.e., in the latter world, states may and do simply disallow migrants from receiving welfare benefits). If there is substantial welfare-induced migration in this metric, the optimal welfare benefit level would be much greater in the no-welfare migration world than in the actual world.

I show theoretically that one can write the approximate percentage increase in the optimal benefit level in the no-welfare migration world, relative to the optimal benefit level in the real world, as the ratio of two effects. The numerator of this ratio is the “welfare migration effect.”

---

98 Mary Jo Bane and David T. Ellwood (1994), Welfare Realities: From Rhetoric to Reform, Cambridge, MA: Harvard University Press. While dated, this book is chock full of stimulating analysis and facts. Of special interest is the lucid discussion of welfare use dynamics in Chapter 2, including the fascinating facts, collected from a single data set, that (i) about a third of current welfare recipients had been receiving welfare for twenty-four consecutive months or less, (ii) half of all completed welfare spells, defined as consecutive time on welfare, were shorter than twenty-four months, (iii) fewer than 15 percent of all current welfare recipients would be on welfare for a total of twenty-four months or less, and (iv) nearly half of all current welfare recipients would be receiving welfare ten years later; see p. 29.

99 Bane and Ellwood (1994), supra note 98, at 52, Table 2.6.
which captures the increase in welfare-related spending that occurs when benefits are increased. The denominator is the “constant-population effect,” which captures the change in welfare spending that occurs because an increase in generosity causes the state to pay more benefits to welfare participants who would have lived in the state regardless of the benefit change. I then use my empirical estimates as inputs to a simulation procedure meant to approximate these welfare migration and constant population effects. My results suggest that optimal welfare benefit levels in Illinois and Wisconsin—a representative medium- and high-benefit state—would be no more than 10% greater if there were no welfare migration. In sum, my results suggest that, even accounting for previously neglected lifecycle considerations, the amount of welfare migration is probably too small to have large effects on optimal state welfare benefit levels, at least for relatively high-benefit states.

B. Summary Thoughts on the Incentive Effects of AFDC

In general, careful studies did tend to show causal effects of AFDC on behavior. As Robert Moffitt concluded in his widely respected review of the literature through the early 1990s, there was “unequivocal evidence of effects on labor supply, participation in the welfare system, and on some aspects of family structure,” though “the importance of these effects is limited in many respects.” Moffitt, Incentive Effects of the U.S. Welfare System, at 56. Given this evidence, the ultimate policy consensus in favor of eliminating AFDC seems was out of any real proportion to the evidence.

AFDC had some negative behavioral effects, to be sure, and AFDC certainly did not eliminate poverty. But how could it, given the maximum benefit levels? In 1996, the official poverty line for a family of three was $1,053 per month. As the final column of Table 1 shows, no state had an AFDC maximum monthly benefit level close to this level. Given the 100 percent statutory benefit reduction rate, the maximum monthly benefit level is an effective ceiling on income. Thus, AFDC could hardly reduce officially measured poverty as a general matter.

Note that Food stamps payments are not included in income for purposes of computing official poverty statistics, so this feature of the programs would not change the conclusion in

100 See Table 6, at 1116.
101 For Mississippi, a very low-benefit state, the results are sensitive to assumptions I discuss in detail in the paper; alternative point estimates suggest that optimal welfare benefit levels might be only 2.1% greater in the absence of welfare migration, but they might also be as great as 36.5% greater; see Table 6, at 1116. Why would Mississippi’s optimal benefit level be a function of welfare migration, given that Mississippi’s low benefit level should operate to repel in-migration? Because when welfare migration does occur, Mississippi can convince some of its poor residents to leave by reducing its welfare benefit level.
103 According to the Poverty Thresholds table at http://www.census.gov/hhes/www/poverty/data/threshld/thresh96.html, the annual poverty threshold was $12,641 for a family with three persons, of whom two are related children. Dividing this annual figure by 12 yields $1,053 per month.
the previous paragraph. But even if food stamps benefits had been counted in measuring official poverty, Figure 7 shows that it would make no difference. The figure plots each state’s maximum combined AFDC and food stamps benefit as a share of the federal poverty line (vertical axis) against the maximum AFDC benefit alone (horizontal axis). It shows that even the combined program benefit in 1996 was below the poverty line—and far below it for many states.

Figure 7: 1996 Monthly Maximum Benefit Levels as Share of Federal Poverty Line

Thus, one could scratch his head bald at the thought that the AFDC program could have reduced, much less eliminated, officially measured poverty. That’s simply not what the program was built to do. That said, AFDC did transfer a nontrivial amount of dollars to poor households, and the failure of the poverty rate to indicate this fact is simply a measurement problem. The poverty rate is just the sum of a dummy variable that equals one for those with income below the threshold and zero for everyone else. Thus, transfers of income across from those above the line to those below the line do not affect the poverty rate unless the transfers are sufficient to raise some people above the line.

In retrospect, one remarkable feature of the micro-econometric literature on welfare policy is just how little attention was paid to the benefits side of the equation. To be sure, it is substantially easier to measure a transfer program’s efficiency costs, in the form of incentive effects, than it is to measure benefits; this is especially true because the benefits often involve highly normative judgments related to distributive justice. Perhaps as a consequence, economists tend to spend most of our time estimating the costs of running welfare programs. But there are alternatives. For example, we could ask and try to answer questions like, how many dollars of

the marginal taxpayer’s would have to be equivalent to an additional dollar in the hands of a low-income person in order for a welfare program’s incentive costs to be worth bearing?

Moreover, some forms of benefit can be measured more directly. Gruber (2000) estimates the effects of AFDC in smoothing consumption following a divorce.\footnote{Jonathan Gruber, “Cash welfare as a consumption smoothing mechanism for divorced mothers,” Journal of Public Economics 75 (2000) 157–182.} He finds that AFDC benefits do not crowd out other means of consumption smoothing (e.g., support from relatives or earnings); see Gruber (2000) at 157. This finding suggests that AFDC was a particularly effective form of consumption insurance, at least against the risks associated with divorce.\footnote{There might be disagreement concerning the desirability of reducing the negative consumption shocks associated with divorce. For example, consider the allegorical Pennsylvania couple Harold and Phyllis whom Charles Murray discusses in his book Losing Ground. Charles A. Murray (1984), Losing Ground: American Social Policy, 1950–1980, Basic Books. Murray discusses whether the couple would marry after Harold gets Phyllis pregnant. Before the welfare policies Murray opposes, he believes Harold would have married Phyllis, and that that would have been a good thing. But after these policies take effect, Murray argues, they will not marry, because it makes no economic sense to do so. Murray criticizes this result as an example of “incentives to fail” that “maximiz[e] short term gains” (the chapter in which he discusses Harold and Phyllis is titled, “Incentives to Fail I: Maximizing Short Term Gains”). In a review of Losing Ground, “How Poor Are the Poor?” The New York Review of Books, May 9, 1985, Christopher Jencks contests Murray’s suggestion that Harold and Phyllis’s marriage would be an unequivocal social good, even granting some of Murray’s premises concerning Harold and Phyllis, like whether Harold would necessarily have Phyllis her before the welfare policies Murray decries. Jencks writes:

Shorn of rhetoric, then, the “empirical” case against the welfare system comes to this. First, high AFDC benefits allow single mothers to set up their own households. Second, high AFDC benefits allow mothers to end bad marriages. Third, high benefits may make divorced mothers more cautious about remarrying. All these “costs” strike me as benefits. It seems unlikely that Murray would agree with that last sentence; thus it seems equally unlikely that he—or those who viewed Losing Ground as a call to action—would approve of AFDC’s role in smoothing divorce-induced negative consumption shocks.} Empirical evidence also suggests that by helping allow mothers of infants time out of the labor force, AFDC increased breastfeeding rates.\footnote{See Steven J. Haider, Alison Jacknowitz and Robert F. Schoeni, “Welfare Work Requirements and Child Well-Being: Evidence from the Effects on Breast-Feeding,” Demography, Vol. 40, No. 3 (Aug., 2003), pp. 479-497, who write that their results show that “if welfare reform had not been adopted, national breast-feeding rates six months after birth would have been 5.5% higher than they were in 2000.”}

Recent evidence also suggests that the introduction of the Food Stamps program led to increased birth weight, especially among those babies born at lower birth weights. Given the close relationship between the food stamps and AFDC programs that I discussed in section II.A.1.d, both in terms of eligibility and in terms of program structure, this finding provides more evidence that income support like that provided by AFDC has real benefits.\footnote{See Douglas Almond, Hilary W. Hoynes, and Diane Whitmore Schanzenbach, “Inside the War on Poverty: The Impact of Food Stamps on Birth Outcomes,” Review of Economics and Statistics, v93, n2, May 2011, pp. 387-403.} So while we do not have much quantifiable evidence on the benefits of the old welfare system, especially compared to the evidence on incentive costs, it is unlikely that that is because there weren’t any such benefits.

C. Welfare Reform: Waivers and PRWORA
Before the wave of waivers and PRWORA, states had very little control over program details. They could set benefit levels. And because they administered their own AFDC programs, they could choose the levels of efficiency, bureaucracy, and, within limits, harassment of participants they wanted to use.\(^{109}\) As mentioned above, AFDC’s structural inflexibility began to give way under various waivers to the AFDC rules. Approved under authority of Section 1115 of the Social Security Act,\(^{110}\) these waivers first were allowed in the Reagan administration and really took off under President Clinton. I discuss the details of programmatic reforms in subsection 1 and then turn in subsection 2 to the empirical evidence on these reforms’ effects. Much of my discussion in section 1 draws from Rebecca M. Blank’s excellent 2002 review in the Journal of Economic Literature, *Evaluating Welfare Reform in the United States*,\(^{111}\) while subsection 2 is drawn almost entirely from Jeffrey Grogger and Lynn A. Karoly’s impressive book, *Welfare Reform: Effects of a Decade of Change*.\(^{112}\)

1. What Welfare Reform Entailed

By the time of PRWORA’s enactment, 27 states had implemented major waivers operating statewide.\(^{113}\) These waivers, which required federal approval, changed many aspects of existing AFDC programs. Under the federal rules, AFDC was an entitlement for anyone who met both federal categorical eligibility conditions and a state’s means test. Thus, depending on the spacing of a woman’s births and her income, she could in principle retain AFDC eligibility for decades (though I am aware of no empirical work suggesting this was a common occurrence).

To illustrate the differences between AFDC and TANF programs that PRWORA allowed, one need only consider Connecticut’s TANF program, itself begun as a waiver called Jobs First. This program included a lifetime time limit of 21 months on assistance.\(^{114}\) Oddly, Jobs First wields this rattan cane alongside the dangling carrot of a liberalized NIT structure, which allows welfare participants to earn substantial incomes while retaining program eligibility.\(^{115}\) Time limits and liberalized benefit reduction rates were only two of the potentially important changes states incorporated into their waiver programs. Other examples include family caps that eliminated the generally meager increase in benefits states had provided following birth of another child (see section II.A.1.c, above); work requirements; substantial sanctions for failing to participate in required work activities; more generous treatment of earned income and child

\(^{109}\) For a theoretical argument showing that bureaucratic hassles and other “ordeals” can be efficient, because they deter applications from less needy people when the level of neediness can’t be observed, see Albert L. Nichols and Richard J. Zeckhauser (1982), “Targeting Transfers through Restrictions on Recipients,” *American Economic Review*, Vol. 72, No. 2 (Papers and Proceedings) (May), pp. 372-377.

\(^{110}\) Codified at 42 U.S.C. 1315.


\(^{112}\) Grogger and Karoly (2005), *supra* note 6.

\(^{113}\) Blank (2002), *supra* note 6, at 1106.


\(^{115}\) I discuss the details further below.
support in benefit determination; more generous child care and Medicaid assistance; and novelty features like required drug testing.\textsuperscript{116}

PRWORA is sometimes thought of as basically changing default rules, so that now there is a presumption in favor of state flexibility.\textsuperscript{117} This is true in some regards, and as I discuss below, state TANF rules certainly do vary widely. On the other hand, the PRWORA regime has its own rigidities. For example, states are required to impose lifetime time limits on TANF receipt of no greater than 60 months,\textsuperscript{118} and PRWORA required that by 2002, at least half of families receiving TANF-funded aid be engaged in qualified work activities.\textsuperscript{119}

States responded to PRWORA by creating programs with the same kinds of diversity seen in the waiver programs. They adjusted benefit reduction rates, created welfare-to-work programs, implemented various sanctions to punish recipients who fail to follow work-activity requirements, and imposed time limits. In addition, they took advantage of the elimination of the entitlement to cash assistance by developing “diversion” programs to encourage applicants not to take up cash welfare in the first place.\textsuperscript{120} Some states also increased work-support subsidies, including for child care, transportation, and various types of wage subsidies.\textsuperscript{121}

2. Evidence on the Effects of Welfare Reform

The literature on welfare reform’s effects is enormous. Grogger & Karoly’s book \textit{Welfare Reform: Effects of a Decade of Change}\textsuperscript{122} has separate chapters on welfare reform’s effects on each of the following topics: welfare use and payments; employment, labor supply and earnings; income and poverty; family structure; and child outcomes. Grogger & Karoly report that most of the reforms considered in the papers they reviewed led to increases in employment.\textsuperscript{123} Some policies, like financial work incentives, increased welfare use. Others, like time limits and

\textsuperscript{116} See Corinne A. Carey, \textit{Crafting a Challenge to the Practice of Drug Testing Welfare Recipients: Federal Welfare Reform and State Response as the Most Recent Chapter in the War on Drugs}, 46 Buff. L. Rev. 281, 297 (1998) (“South Carolina, the recipient of one such waiver, became the first state to make drug testing part of its welfare program.”).

\textsuperscript{117} See, e.g., Rebecca M. Blank, \textit{Evaluating Welfare Reform in the United States}, 40 Journal of Economic Literature 1105, 1106 (stating that the block granting of TANF “essentially removed almost all federal eligibility and payment rules, giving states much greater discretion in designing their own cash public assistance programs. This also eliminated a federal entitlement to cash assistance. States could choose which families they supported.”)

\textsuperscript{118} See, e.g., \textit{id.}; note also that “[s]tates could exempt up to 20 percent of their caseload from this limit, could set shorter time limits if they chose, or could continue funding assistance to families entirely out of state funds after sixty months.” \textit{id.}

\textsuperscript{119} \textit{Id.} Note that PRWORA treated caseload reductions and work as perfect substitutes for this purpose: caseload reductions were credited toward the work activity requirement, so that “a state that reduced its caseload by 50 percent would meet its work requirement, regardless of how many current or former recipients were actually employed.” \textit{id.}

\textsuperscript{120} According to Blank, 30 states used some sort of diversion program. \textit{id.}, at 1114.

\textsuperscript{121} \textit{id.}

\textsuperscript{122} Grogger and Karoly (2005), \textit{supra} note 6.

\textsuperscript{123} See Figure 10.1 and text at 231-233.
sanctions for non-compliance with program rules, reduced welfare use.\textsuperscript{124} Overall, the studies GK reviewed vary as to the effects of reform on income.\textsuperscript{125}

None of the studies GK reviewed find evidence that welfare reform increased poverty, while several of the estimates they reviewed suggest that reform programs actually reduced poverty.\textsuperscript{126} In principle this set of results is in line with the labor supply theory discussed above. Since maximum benefit levels were substantially less than the poverty line, even wholesale elimination of welfare eligibility wouldn’t increase the official poverty rate. On the other hand, if there are large entry effects, then providing AFDC could have caused increases in the officially measured poverty rate by causing women who would earn more than the poverty threshold in the absence of the AFDC program to forgo paid work in favor of the combination of lower consumption and zero/fewer work hours. Thus, it is theoretically possible that eliminating AFDC altogether might reduce the number of women whose income falls below the official poverty line. To the extent that welfare reform eliminated welfare eligibility for some women, it fits within this analysis.

But the evidence that welfare reform reduced poverty tends to come from studies evaluating programs that involved generous financial incentives to work,\textsuperscript{127} which are present only in some states. Thus I am dubious of claims that reductions in the poverty rate attributable to welfare reform were the per se result of eliminating AFDC’s labor supply disincentive. Rather, at least some of these reductions in poverty were due to the more active role that government played in subsidizing those who work. This fact should be no surprise to anyone familiar with the literature on federal (and state) EITC programs.\textsuperscript{128}

In general, GK found mixed results on marriage and childbearing, with some studies showing effects in each direction and relatively few showing statistical significance.

Finally, note that much of the evidence reviewed by GK—whether on poverty or other outcomes—comes from state-time periods when time limits did not yet apply.\textsuperscript{129} Once the time limit binds for a given woman, her state’s welfare program essentially has been eliminated for her.\textsuperscript{130} At this point, any positive effects on work, income and other outcomes that result from more active state involvement in the woman’s case can be expected to disappear, or at the least attenuate substantially. As a result, it is unclear whether reforms built on states’ active labor

\textsuperscript{124} See Figure 10.2 at 232.
\textsuperscript{125} Id.
\textsuperscript{126} See Figure 10.2 at 232 and Figure 10.5 at 247.
\textsuperscript{127} See text at 234.
\textsuperscript{129} See text at 247 (noting the fact that results in their Figure 10.5 “represent the effects of reform as a bundle during the pre-time limit period. Post-time limit evidence is very limited, and most studies summarized in this figure cover time periods prior to when recipients could have exhausted their benefits.”). Also, note that only some of the randomized experiments from which GK’s evidence on detailed aspects of welfare programs is drawn involved time limits.
\textsuperscript{130} If states do not share data, then as many have noted, time limited recipients could migrate to avoid time limits; see Gelbach (2004), supra note 97, and, e.g., Hal W. Snarr and Mark L. Burkey, “A Preliminary Investigation of Welfare Migration Induced by Time Limits,” Journal of Regional Analysis and Policy, 36(2): pp. 124-139.
market involvement will persist in the long run. Unfortunately, we do not have the kind of variation in programs necessary to distinguish between short and long run effects.

GK conclude that the evidence “is generally consistent with the outcomes about which” their (quite standard) behavioral model “makes the clearest predictions.” They note, though, that a number of outcomes—including poverty, marriage, childbearing, and child well-being—have been understudied. In addition, they point out that most of what we know about the causal effects of welfare reform comes from relatively short follow-up periods, whereas, as I note just above, an assessment of whether welfare reform “worked” or didn’t really involves long run impacts. While I have the utmost respect for the job that GK did describing and assessing an overwhelming amount of research, I have argued elsewhere that one drawback of their book is that it does not attack the question of “just how much welfare reform ‘worked,’ and how much of the work was due to other factors (including those that may have interacted with welfare reform, as noted by GK[]).” This is not an easy question, though I do take a partial crack at it below.

For other discussions and reviews of the evidence, as well as some research conducted after GK’s publication I refer readers to the following sources:


III. Welfare Reform and the Laboratories of Democracy Analogy

---

131 This discussion recalls the point from section II.A.2, concerning the two programs—one providing transitory help and one providing more or less permanent assistance—implicitly embedded in AFDC. Almost by definition, the women most likely to be in need of long run labor market help are those served by the long run AFDC program. Thus, these are the women most likely to benefit from welfare reform’s carrots—which PRWORA eliminates when these women reach the time limit.

132 There is plenty of individual-level data, to be sure. But as I argue in section II below, cross-state variation in state program rules no longer is the simple source of identifying variation that it was in the pre-welfare reform era. That leaves researchers to use within-state designs, but without random assignment experiments, it is difficult to separately identify the effects of infrequent policy changes—like introduction of a state’s time limit together with active labor market involvement—from effects of unrelated labor market variables, like labor demand. Finally, random assignment studies tend to be relatively short, with at most a few years of follow up, which makes it difficult to use them to assess effects over more than the medium-to-long run.

133 GK at 248.

134 Id.

135 Id.

In this section, I discuss welfare reform through the lens of Justice Brandeis’s laboratories of democracy metaphor. Without variation of one type or another, of course, nothing can be learned about the empirical world; but laboratories are only as good as the experiments their operators conduct. Early in the welfare reform process, the federal requirement that states conduct evaluation studies led to a number of high-quality evaluations of welfare reform programs. A number of these involved random assignment, and public use data are available for several. Even the best of these experiments provide somewhat limited guidance for future policy choices, as most of them stirred together multiple reforms, allowing experimental evaluation of only the resulting soup. But the multi-faceted, nature of post-PRWORA 1990s devolution was a considerably worse development from a laboratories perspective. States introduced dozens of reforms at once, effectively destroying the capacity of even quite savvy researchers to extract clear signals from the data about the details of state welfare reform policies.

As I discuss in Part II.A above, there was relatively little structural variation in the AFDC system, with a consensus among empirical economists being that benefit level variation was far and away the most important source of cross-state program variation. But by the time the waiver-era and PRWORA dust settled, there were essentially 51 different welfare programs. To illustrate this point, consider the heroic efforts of the Urban Institute’s “Assessing the New Federalism” (UI-ANF) project. This project produced hundreds of studies between the mid-1990s and 2007, focused on the design, implementation, and effects of the waivers and PRWORA.

One part of the UI-ANF project was the “Welfare Rules Database” (WRD), wherein the Urban Institute undertook the herculean task of collecting and categorizing the various aspects of state welfare policies. There are thirty categories of variables included in the WRD! In alphabetical order, these categories are:

Activities exemptions, Activities requirements, Activities sanctions, Asset test, Benefit computation, Child support, Child support sanctions, Components, Contracts and agreements, Countable income, Deemed income, Diversion, Dollar amounts, Earned income disregards, Eligibility by number/type parents, Eligibility of individual family members, Eligibility of pregnant women, Eligibility of two-parent families, Eligibility of units headed by minor parents, Family caps, Immunization and health screening policies, In kind income, Inclusion of non-citizens in the unit, Income and assets of children, Income eligibility tests, Minor parent activities and bonuses, School

---

137 For example, MDRC posts a list of available files at http://www.mdrc.org/pudf_available.htm, with instructions on how to request them at http://www.mdrc.org/pudf_howto.htm.
138 A notable exception is MDRC’s evaluation of Minnesota’s Family Investment Program, or MFIP. This random assignment evaluation included three study groups: the AFDC control group; the main MFIP program group, which faced work requirements and was provided financial incentives to work; and the MFIP-Incentives Only group, which received financial incentives to work but did not face work requirements. See Lisa A. Gennetian, Cynthia Miller, and Jared Smith, Turning Welfare into a Work Support: Six-Year Impacts on Parents and Children from the Minnesota Family Investment Program, MDRC, July 2005, available at http://www.mdrc.org/publications/411/full.pdf.
139 As of February 12, 2012, the UI-ANF website lists 362 reports; see http://www.urban.org/center/anf/Reports.cfm, with the most recent dated August 3, 2007.
policies for dependent children, Time limits, Transitional benefits, and Treatment of Additional Adults in Household.

And remember: each of these thirty categories includes multiple variables that characterize a state’s welfare program. As a result, the WRD contains hundreds of variables.\textsuperscript{140} And even welfare reform didn’t change the fact that there are only 51 state-level jurisdictions. In other words, our system of state laboratories imposed more experimental treatments than there were experimental units.

Researchers interested in estimating causal effects of policy on behavior and outcomes have responded to this challenge in four primary ways. The first response is to use data from state-specific waiver evaluation studies, several of which used randomization of program assignment to assign welfare participants and applicants to either the old AFDC program or the state’s waiver program. Examples of such work includes various official reports that professional program evaluation groups such as ABT, Mathematica, MDRC, and the Urban Institute produced under contract with states.\textsuperscript{141} Other examples include journal articles using data from these experiments.\textsuperscript{142}

A second response to the too-much-variation problem is to apply non-experimental multivariate methods to repeated observations on data from multiple states, with simplifying assumptions used to choose the treatment variables considered. One such approach is to compare state-year cells in which some form of welfare reform—waiver or TANF—is in effect to state-year cells in which no reform operates.\textsuperscript{143} The resulting coefficient estimates are then interpreted

\textsuperscript{140}While it is not obvious from the WRD’s website how many variables the database contains, a simple lower bound can be discerned by searching the variable descriptions for those containing the word “the.” This query (at http://anfdata.urban.org/wrd/query/WRDSearch_result.cfm?desc=the) returned 576 variables when I conducted it on November 19, 2011.


as some sort of average impact of implementing the given reform (waiver or TANF) among implementing states. Because TANF implementation occurred over a brief, 16-month window, after which all states had TANF in effect, there is relatively little variation with which to identify its effects in this way. Moreover, if these effects vary over time, then all one can hope to identify concerning TANF are its short-run implementation effects.\(^{144}\)

Another simplifying approach is to try to classify various welfare-plan characteristics into specific variables thought to be substantively interesting or important.\(^{145}\) In one of the more ambitious such studies, Urban Institute researchers Signe-Mary McKernan and Caroline Ratcliffe examined the empirical relationship between poverty and 19 welfare policy variables.\(^{146}\) Of these variables, the authors hypothesize that one should have an unambiguously positive effect on poverty and five should have an unambiguously negative effect, with the theoretical effects of the other 13 being ambiguous. Summing up their results, McKernan and Ratcliffe write that “[o]ur findings are generally consistent with our hypotheses,” and I have no quarrel with that conclusion. Nor do I mean to criticize them for their serious and substantial efforts. Nonetheless, it is difficult to know quite what policy lessons to draw from a study that includes as regressors 19 inter-related policy variables, even as it excludes numerous others. One simply can’t tell whether these are the “right” variables, either in the sense that they are the most important from a policy perspective or in the sense that there are no important omitted policy variables that are correlated with, and thus proxied by, the included ones.

A third response to the too-much-variation problem is altogether to ignore the issues of cross-state policy variation and confounding factors, instead just comparing outcomes of interest before and after welfare reform occurred. This is the approach taken in an influential paper by Bruce Meyer and James Sullivan.\(^{147}\) Meyer and Sullivan compare various measures of income and consumption before and after the welfare reforms of the 1990s,\(^{148}\) finding no evidence of increases in material deprivation among single mothers.\(^{149}\) They do find drops in a measure of well-being that places positive value on time spent in non-paid work activities, which is not surprising given the large increase in market work among single mothers over their study period.

---


\(^{149}\) Meyer and Sullivan (2008), at 2237.
A fourth approach is to try to use state program rules together with individual characteristics to measure the budget set facing each survey subject independently. One example of this approach is Fang and Keane (2004),\textsuperscript{150} who painstakingly construct individual-specific budget constraints from the many dimensions along which state policies vary over time. They then exclude state and year dummies from their analysis, which amounts to assuming that their budget constraint measures capture every systematically important determinant of behavior across states and over time. While they regard this assumption as a feature, not everyone need agree.\textsuperscript{151}

Francesca Mazzolari focuses primarily on time limits, rather than trying to code the entire budget constraint.\textsuperscript{152} She uses data covering 1989-2003 from overlapping panels of the Survey of Income and Program Participation. Using panel data allows her to measure the remaining welfare program eligibility for each woman in her sample, rather than using only variables that quantify overall state policy.\textsuperscript{153} Mazzolari finds that by 2003, time limits per se had reduced national caseloads by 25 percent.\textsuperscript{154} Intriguingly, only about 5 of the 25 percent are explained by behavioral responses,\textsuperscript{155} understood as women reducing their use of welfare in advance of the time limit in order to preserve future eligibility. The remaining 20 percentage points are the result of the mechanical effect that time limits eliminate eligibility once they bind. Thus, Mazzolari’s paper suggests that very little of the large drop in welfare caseloads can be attributed to behavioral, as opposed to mechanical, changes related to time limits.

A final possible response to the too-much-variation problem is to simply stop trying to study the causal effects of welfare reform.\textsuperscript{156} Judging only by the numbers of papers I’ve received to referee and have seen on conference agendas—hardly a scientific measure, I readily concede—there seems to have been less focus on “traditional” cash welfare policy, at least in the economics research community, than there was either before or during the wave of 1990s reform. The attention paid to poverty and the problems of low income people more broadly is focused now much more on other policy avenues like expansion of health coverage and EITC programs at the state or federal level. Of course, even if I am right about this drop in focus, perhaps it is simply the natural result of a public policy debate that has largely concluded. On this theory, we


\textsuperscript{151} For an argument that this assumption is decidedly a bug, see Rebecca Blank’s comments in the “Comments and Discussion” section of Fang and Keane (2004), starting especially at page 97.


\textsuperscript{153} Such direct measurement is important because a woman who has used welfare for 59 months in a state with a 60-month time limit faces very different incentives from an otherwise-identical woman in the same state who has never used welfare.

\textsuperscript{154} Mazzolari (2007), at 616.

\textsuperscript{155} Mazzolari (2007), at 616.

\textsuperscript{156} As a cheap-kill example (and one that I do not intend as a criticism of the Urban Institute or its many outstanding scholars and researchers), I note that the Urban Institute changed the name of the Assessing New Federalism project to the “Low-Income Working Families” project, whose stated raison d’etre is no longer to “assess” welfare reform but rather to “appl[y] rigorous research methods to track families over time and to analyze the risks these families face.” See http://www.urban.org/center/lwf/index.cfm.
had a national conversation about how to get rid of traditional welfare policy, we pretty much got rid of it, and now there just isn’t much of interest to discuss.\footnote{157}

But this seems too simple to me, because researchers do continue to study Medicaid, the EITC, unemployment compensation, various programs operating under the aegis of the Social Security Administration, and other social policies that have substantial effects on those with low incomes.\footnote{158} And if nothing else, graduate students need jobs and assistant professors need tenure. As such, I am convinced that if there were good variation to use, researchers would use it. Of course, the provision of good long run research opportunities was hardly the point of welfare reform. My point is simply that it shows. Welfare and welfare reform constitute a policy field where the evidence that exists largely followed the reforms, rather than providing the empirical inputs to a policy production function. And the reforms themselves have largely blocked information gathering. From a policy-design perspective, the story of welfare policy devolution is a cautionary tale concerning what can be learned from a simultaneous explosion of state-level experimentation. By my lights, anyway, welfare reform has been no exemplar of Justice Brandeis’s laboratory.

\section*{IV. Did Welfare Reform “Work”? Experimental Evidence from Connecticut}

Perhaps the most frequently discussed outcome variables in both academic and policymaking circles are welfare caseloads and employment rates of single mothers. As Rebecca Blank has written, “[i]t remains very difficult in any of this literature to separate economic and policy effects. . . . The combination of extremely strong economic growth with rapid policy change means that both effects were reinforcing each other.”\footnote{159} Thus, Blank wrote, this important question “remains a difficult [one] to answer in an entirely credible manner.”\footnote{160}

In this part of the chapter, I use experimental data from Connecticut, which conducted a high quality random assignment study of its Jobs First welfare reform.\footnote{161} Because it included arguably the sweetest carrot and biggest stick of reforms nationwide, I focus primarily on Connecticut’s Jobs First evaluation. The evaluation involved a random assignment experiment of the state’s waiver program, which subsequently formed the basis of Connecticut’s TANF program. In section A of this part of the chapter, I review the facts and literature on caseloads and the economy. I then discuss the Jobs First program, and its differences from AFDC, in section B. In section C, I present results from Connecticut on welfare participation, employment,

\begin{footnotes}
\footnote{157} Of course, former recipients of cash assistance might feel differently. But my sense was and continues to be that relatively little of the welfare policy debates focused on the desires of the recipients.
\footnote{158} Indeed, Janet Currie has written a valuable book making the important point that these programs together constitute a safety net that is critical to the well being of the disadvantaged—and that itself may come under the sort of political attack that felled AFDC. See Janet M. Currie, The Invisible Safety Net: Protecting the Nation’s Poor Children and Families, Princeton University Press (2008).
\footnote{159} Blank (2002), supra note 6, at 1136.
\footnote{160} Id.
\footnote{161} This section draws on an earlier working paper of mine, Jonah B. Gelbach, Much Ado About Relatively Little? Experimental Evidence on Welfare Reform, Caseloads, and Employment, typescript, December 8, 2004, available at http://gelbach.law.yale.edu/papers/caseloads-paper-dec-8-04.pdf. As that paper is unpublished and I have no plans to submit it for publication anywhere, I will use some of the text verbatim in the present chapter.
\end{footnotes}
and income. I report analogous results from several other states’ random assignment studies below.

A. Background on Caseloads and the Economy

As illustrated in Figure 5, the average monthly welfare caseload fell enormously—by more than half —after peaking at about five million cases in 1994. At the same time, employment rates for single mothers rose by a remarkable 10 percentage points, from about 68 to about 78 percent, between 1994 and 2000.

Many popular commentators and policymakers have concluded that welfare reform caused these changes. As I discuss in Part II.C, above, economic theory suggests that most elements of reform should reduce caseloads and increase employment (though more generous earnings disregards should actually increase caseloads, other things equal). However, I have noted that welfare reform wasn't the only change over the mid- to late-1990s. There were significant expansions in the EITC. And the traditional link between Medicaid coverage of children and AFDC participation was purposely broken, allowing many children to retain public health insurance even if their mothers left welfare. Such changes should also have caused reductions in welfare participation and increases in employment, as some literature suggests they did.

But the 800-pound gorilla in this literature is the remarkable economic boom that occurred throughout the latter part of the 1990s. The dashed line in the left panel of Figure 8 plots the national caseload and national unemployment rate over the period between 1993 and 2000. The figure shows that both variables fell steadily over this period. Of course, the unemployment rate is just one convenient example; virtually every measure of labor market conditions improved radically over this period. Without further evidence, one cannot rule out the possibility that the caseload and employment changes among the welfare population were due entirely to welfare reform, entirely to the economy, or (more likely) due to some combination of these factors. As noted above, numerous authors attempted to sift the data to answer this question. The share of the caseload drop causally attributed to welfare reform in this literature ranged from a low of zero (or even a negative share, i.e., suggesting that welfare reform may have increased caseloads) to a high of about half.

---

162 All unemployment rates used in this section were downloaded from the Bureau of Labor Statistics (BLS) website and are seasonally adjusted by the BLS.
164 For example, results in James P. Ziliak et al. (2000), supra note 145, imply that two-thirds of the caseload drop is due to macroeconomic improvements, and that “in the absence of other influences, welfare reform would not have led to any decrease in aggregate caseloads,” at 572, though they find that some specific types of waivers do reduce
The share of the caseload drop attributed to the economy ranges from zero to as much as two-thirds. In Rebecca Blank’s words, “These estimates range widely across studies, and are highly affected by the years over which they are estimated and by the specification.” As noted above, Blank concludes that “Despite evaluation and methodological problems, the bulk of the research literature suggests that the policy changes over the 1990s were important to the rapid caseload decline.”


---

165 Blank (2002, JEL), at 1135.
decline before 1998 can be explained by improvements in economic conditions. These two papers are notable not only because their findings suggest the possibility that welfare reform effects were limited, but also because they are based on internally consistent models of flow data. By contrast, most of the rest of the literature focuses on reduced form models of stocks, and Klerman and Haider show that those models are necessarily mis-specified. Both of these papers emphasize that California's experience over the period before welfare reform may not have been representative of other states' experiences. For example, Klerman and Haider point out that “in comparing California to the rest of the nation, the 1990s recession was deeper, welfare reforms were passed later, and the reforms that were passed were weaker. Thus, the role for the economy may be greater in California than in other states.”

All of these papers—and most if not all of the literature seeking to disentangle policy from economic effects—use non-experimental data. These data characterize either individual employment and welfare use decisions or state-level welfare participation rates, employment rates, or caseload levels. Thus, the papers provide valid estimates of the contributions of policy and economic changes only when strong statistical assumptions are satisfied. A random assignment experiment can avoid the need for these assumptions because women in both the treatment and control groups face identical economic conditions: the only reason the welfare programs facing these women are different is the randomized assignment of the women's effective program. Thus we can safely attribute differences in outcomes to the program itself, rather than other factors changing at the same time.

---

168 Klerman and Haider (2004), supra note 166.
169 Klerman and Haider (2004), supra note 166, at 883 (citations omitted).
170 In co-authored work on welfare reform's effects on living arrangements, marriage and divorce, and health variables, I have made similar assumptions; see Bitler, Gelbach, Hoynes, and Zavodny (2004), supra note 143; Bitler, Gelbach and Hoynes (2005), supra note 143; and Bitler, Gelbach, and Hoynes (2006), supra note 143.
B. Connecticut’s Jobs First Program and Evaluation

Jobs First began operating in January 1996 as a waiver from AFDC rules, and Connecticut subsequently adopted it as the state’s TANF program. For present purposes, the program had two key features. First, as noted above, the program had a 21-month lifetime time limit, though participants who complied with other program requirements and had sufficiently low income could receive 6-month extensions; others could receive exemptions with good cause.

Second, Jobs First allowed a 100 percent disregard of earned income up to the poverty line. To understand this provision, recall the modified NIT structure discussed in part II.A above, and ignore the interaction between the food stamps and welfare programs. For a woman with no non-labor income, a welfare program that disregards D dollars in earned income will pay a benefit equal to whatever is greater, zero or

\[ P \equiv G - t \times \max\{0, Y - D\}, \]

so that benefits are reduced only if the participant earns more than D dollars.

Jobs First has the interesting feature that, in principle, the entire benefit payment is lost if the participant earns more than the disregard, D. This means that the benefit reduction rate t is locally infinite. For a less technical understanding of the program, consider Figure 9, which I have copied directly from Marianne P. Bitler, Jonah B. Gelbach, and Hilary W. Hoynes, What Mean Impacts Miss: Distributional Effects of Welfare Reform Experiments. In the figure, the AFDC program is represented by the line connecting points A, C, and B (see Figure 2 above for discussion), while Jobs First is represented by the line connecting A and F. To understand the participation and labor supply effects of moving from AFDC to Jobs First, it helps to imagine that the same woman is trying to decide on her preferred labor supply and welfare participation separately under each program.

If the woman would participate on welfare when the welfare program she faces is AFDC, she either locates at point A or on the AFDC budget segment at some point between A and B, with point C being a generic example of this second location decision. Some women who would locate at A when they faced AFDC would also locate there when they faced Jobs First, while others would enter the labor force while still receiving in welfare, due to the increased return

---


173 It also made work activity mandatory for all women but those with a child younger than one year old; provided sanctions, which meant that the state would reduce benefit payments for violations of program rules; implemented a partial family cap; increased child care assistance; increased Medicaid assistance; and streamlined the rules governing how welfare recipients would receive whatever child support was paid in their children’s behalf. For details on these and other aspects of the Jobs First program, see Dan Bloom, Susan Scrivener, Charles Michalopoulos, Pamela Morris, Richard Hendra, Diana Adams-Ciardullo, Johanna Walter, with Wanda Vargas, Jobs First: Final Report on Connecticut’s Welfare Reform Initiative, at 4-12, Manpower Demonstration Research Corporation, February 2002, available http://www.mdrc.org/publications/90/full.pdf.

from combining work and welfare. Switching from AFDC to Jobs First will also cause a woman who would locate at a point like C to continue to combine work and welfare, so that she locates between A and F. Under the assumption of convex indifference curves, and that leisure is a normal good, the woman would increase work hours, i.e., wind up to the left of point C.

**Figure 9: AFDC and Jobs First Budget Sets**

Now consider a woman who would locate at point D under AFDC assignment. This woman would not participate in welfare if she faced AFDC program rules. However, if we assume that leisure is a normal good, then when she faced Jobs First program rules, this woman would participate in welfare and reduce her work hours, locating on the line connecting A and F, somewhere to the right of D. Next, consider a woman who would locate at a point like E under AFDC assignment. This point involves consumption below 100 percent of the federal poverty line and entails more work than is required to reach consumption equal to the poverty line under Jobs First. That is, the woman can increase consumption and reduce work hours by moving from E to F if she is assigned to Jobs First instead; economic theory predicts she will take this opportunity. Finally, a woman who would locate at point H under AFDC assignment might either stay there under Jobs First assignment or move to point F, which entails both less consumption and fewer work hours. Which she does will depend on her willingness to exchange labor for consumption.

Putting all of this together yields an interesting and complex blend of predictions. First, switching to Jobs First will have no impact on hours worked or earnings for some of the women who would locate at point A and points like H if AFDC were in effect. Second, switching to Jobs First would increase hours worked and earnings for the rest of the women who would locate at point A under AFDC assignment, as well as those who would locate at points like C. Meanwhile,
switching to Jobs First would reduce hours worked and earnings for the rest of the women who would locate at point H under AFDC assignment, as well as all women who would locate at points like D and E. As for welfare participation, it will remain unchanged for women at point A and points like C under AFDC assignment, as well as some at points like H, but it will increase for those at points like D and E, and the rest of those at points like H.

Judging by Figure 9, then, a switch in program rules from AFDC to Jobs First would increase hours worked and earnings for some women, reduce these variables for others, and leave them unchanged for still others. The switch would cause some women to enter welfare participation, but it would not cause anyone to exit welfare. Of course, once the time limit became binding under the Jobs First program, there would be no available welfare program at all, and at that point welfare participation necessarily would fall.\textsuperscript{175} My previous co-authored work shows that the basic predictions of static labor supply theory, augmented by observations related to the time limit, appear to be borne out by data from the Jobs First evaluation.\textsuperscript{176}

In order to evaluate the performance of the Jobs First program, the state hired the Manpower Demonstration Research Corporation, now known as MDRC, to run a random assignment experiment. Details of the experimental evaluation appear in MDRC’s final report,\textsuperscript{177} but it will help to provide some basic facts here. Random assignment occurred between January 1996 and February 1997 in state welfare offices in New Haven and in Manchester.\textsuperscript{178} Both new applicants and incumbent recipients in these offices were included in the experiment, with individuals assigned to each program with probability one-half. A total of 4,803 single-parent cases were included in the experiment,\textsuperscript{179} of which 2,396 subsequently faced Jobs First program rules and the other 2,407 faced AFDC program rules. Experimental cases were followed for four years, with key post-random assignment data on monthly welfare benefit (including food stamps) payments and quarterly earnings coming from state administrative records. Some pre-random assignment data on benefit payments and earnings are available as well, and a baseline intake file includes information on some demographic variables. MDRC also did a detailed follow-up survey three years into the evaluation. Data from the evaluation have been used in numerous

\textsuperscript{175} This discussion shows that Jobs First is beset with what one might call multiple policy disorder: it encourages women to combine welfare and work, only to evict them from the welfare premises via time limits when they take the hint. This point has been made elsewhere. For example: “combining [the disregard expansion and time limit] complicates the program message: It is difficult to urge recipients both to leave welfare quickly in order to ‘bank’ their available months and to take advantage of a disregard by combining work and welfare.” See Bloom, D., Kemple, J. J., Morris, P., Scrivener, S., Verma, N. & Hendra, R., \textit{The Family Transition Program: Final Report on Florida’s Initial Time-Limited Welfare Program}, Manpower Demonstration Research Corporation, New York, NY (2000).


\textsuperscript{178} The state implemented the Jobs First program rules for cases served through all other offices. Thus, beginning in January 1996, the experimental control group members were the only ones still facing AFDC program rules.

\textsuperscript{179} Of these, 30 cases are missing some data from the public use data set. Where appropriate, I exclude these cases in the analysis below.
papers,\textsuperscript{180} and public use files are available from MDRC for anyone who follows a simple request process.\textsuperscript{181}

Look again at panel (b) of Figure 8, which shows the trend in Connecticut’s statewide welfare caseload and unemployment rate during the experimental follow-up period. It shows that both variables fell steadily over this period, as with the nationwide time series for these variables. Thus, Connecticut’s experience is broadly in line with the nationwide experience over the period of interest.

It also will be useful to compare the Jobs First experimental sample to the statewide and nationwide populations of welfare recipients. The first column of Table 4 provides some basic summary statistics concerning the Jobs First experimental sample. The second and third columns provide information on the same variables for the Connecticut statewide and nationwide caseloads, which are based on my tabulations of 1993-1997 AFDC Quality Control data.\textsuperscript{182} Compared to those in the Jobs First experimental population, the statewide Connecticut caseload is more Hispanic, less black, less white, more likely to be dropouts, and living in households with larger cases. In general, then, the characteristics are reasonably similar across columns (leaving aside the fact that by construction, all Quality Control observations are current welfare recipients, whereas roughly 40 percent of Jobs First experiment observations are new applicants).


\textsuperscript{181} See \textit{supra} note 137.

\textsuperscript{182} The AFDC-QC system was implemented to “ensure that the program’s funds were going to people who were eligible, and to determine whether those who were eligible were receiving the amount to which they are entitled” (this quotation was taken from the Urban Institute’s AFDC Data Archive documentation page, available at \url{http://afdc.urban.org/AFDocumentation.html}).
Table 4: Characteristics of Jobs First Experimental Sample Compared to Overall Connecticut and Nationwide Caseloads

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>Connecticut</th>
<th>National</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Jobs First Experiment</td>
<td>Quality Control</td>
</tr>
<tr>
<td>White</td>
<td>0.36</td>
<td>0.33</td>
</tr>
<tr>
<td>Black</td>
<td>0.37</td>
<td>0.30</td>
</tr>
<tr>
<td>Hispanic</td>
<td>0.21</td>
<td>0.35</td>
</tr>
<tr>
<td>Dropout</td>
<td>0.34</td>
<td>0.39</td>
</tr>
<tr>
<td>HS Diploma</td>
<td>0.56</td>
<td>0.51</td>
</tr>
<tr>
<td>More than high school diploma</td>
<td>0.06</td>
<td>0.10</td>
</tr>
<tr>
<td>At least 4 persons in case</td>
<td>0.22</td>
<td>0.30</td>
</tr>
<tr>
<td>Head younger than 25</td>
<td>0.29</td>
<td>0.29</td>
</tr>
<tr>
<td>Head aged 25-34</td>
<td>0.41</td>
<td>0.45</td>
</tr>
<tr>
<td>Head aged 35 or older</td>
<td>0.29</td>
<td>0.27</td>
</tr>
<tr>
<td>Recipient at time of random assignment</td>
<td>0.61</td>
<td>1.00</td>
</tr>
<tr>
<td>Current spell length 24 months or more</td>
<td>0.60</td>
<td>0.53</td>
</tr>
<tr>
<td>Number of observations</td>
<td>4,773</td>
<td>4,227</td>
</tr>
</tbody>
</table>

C. Changes in Welfare Participation and Employment

Figure 10 plots the shares of incumbent recipients included in the Jobs First treatment and AFDC control groups who received cash assistance in each of the 48 months after random assignment, until month 22 after random assignment, the first month when the Jobs First time limit could bind for anyone in the experiment, the welfare participation rate was actually greater among Jobs First than AFDC group members. This is to be expected, given the discussion in part IV.B. In month 22, the Jobs First participation rate declines sharply, by about 11 points, which is strong evidence that the time limit was enforced for at least some women. For the rest of the period represented in Figure 10, the Jobs First participation rate is below the AFDC participation rate.

183 A very similar figure appears in the middle of Figure 3 of Bloom, Scrivener, Michalopoulos, Morris, Hendra, Adams-Ciardullo, Walter, and Vargas, supra note 114.
184 I use the first month after random assignment as the baseline. I would prefer to use the month of random assignment, of course, since the first month after random assignment will include some treatment effect (which explains why welfare participation is several percentage points greater for the Jobs First than the AFDC group—86.9 compared to 80.6 percent). However, the administrative data show that only 35 percent of new applicants in the Jobs First group, and 32 percent of those in the AFDC group, have positive cash assistance income in the month of random assignment. By the first month after random assignment, this share leaps 40 percentage points for the Jobs First group and 35 points for the AFDC group. This effect is clearly an artifact of delay in the dating of welfare payments related to approved applications, because welfare participation rates remain high in succeeding months. Thus the most reasonable approach is to use the first month after random assignment.
The first row and first column of Table 5 shows that at baseline, 86.9 percent of Jobs First group members received cash assistance. The second column shows that this share fell by 69.7 percentage points over the 48-month follow-up period. But not all of this decline was caused by Jobs First—some of it would have happened even under AFDC due to the normal factors that induced women to leave welfare over time. Indeed, for the AFDC group, the welfare participation rate, which was 80.6 percent at random assignment, fell 55.7 percentage points over the 48-month follow-up period. Thus, the impact of the Jobs First program on the 48-month change in welfare participation rates was to reduce welfare participation by 14.0 percentage points (69.7 minus 55.7). This impact constitutes a 20 percent share of the Jobs First group’s 69.7 percentage-point change.
Table 5: Key Results on Welfare Participation from Jobs First Evaluation

<table>
<thead>
<tr>
<th></th>
<th>Jobs First Group</th>
<th>AFDC Group</th>
<th>Difference in</th>
<th>Share Estimate</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Baseline‡</td>
<td>Change</td>
<td>Baseline‡</td>
<td>Change</td>
</tr>
<tr>
<td>Full sample</td>
<td>86.9</td>
<td>-69.7*</td>
<td>80.6</td>
<td>-55.7*</td>
</tr>
<tr>
<td>Recipients only</td>
<td>95.6</td>
<td>-75.8*</td>
<td>94.2</td>
<td>-63.6*</td>
</tr>
</tbody>
</table>

‡ Baseline is the first month after random assignment for the full sample and the month of random assignment for the recipients-only sample.
* Differences are statistically significant with p values equal to 0 to three digits.

“Share Estimate” is the ratio of “Difference in Difference” estimate to Jobs First Group “Change” column.

The second row of Table 5 repeats this analysis for the subset of experimental subjects who were incumbent recipients at the time of random assignment (that is, this row excludes new applicants). This group is of particular interest because the set of incumbent recipients at any point in time includes a larger share of those likely to spend a longer overall amount of time on welfare than does the set of new applicants. The results show that the Jobs First group’s welfare participation rate fell by 12.2 percentage points more than the AFDC group’s welfare participation rate did. Relative to the Jobs First group’s drop of 75.8 percentage points, this amounts to only a 16 percent reduction.

Together, these results show that at most a fifth of the drop in welfare participation among those experiencing the Jobs First program is the causal result of welfare reform itself. This is a startling finding given the widely held view among policy makers, commentators, and some scholars that welfare reform was an important driver of the drop in caseloads. Yet here we have a random assignment experiment showing that 80 percent or more of the drop was caused by something else—perhaps the strong labor market, the EITC, or, possibly, the fact that AFDC participation was not a permanent condition for very many women who did use welfare.

One drawback of my approach in this analysis is that it captures only effects of Jobs First on program exit. There likely were also entry effects, i.e., changes in welfare participation among those who were not part of the group of women represented by the experimental study group. The steady state welfare participation rate under any set of program rules depends on both the probability that an incumbent recipient will exit and the probability that a non-recipient will enter welfare. If Jobs First had a larger effect on program entry than it had on program exit, then my one-fifth results are underestimates of reform’s overall relative contribution to the caseload drop.

---

185 For the group of incumbent recipients, the data problem discussed in footnote 184 does not exist. Thus I use the month of random assignment as the baseline for this group.
186 For more on spell dynamics and the composition of cross-sections of current recipients, see, e.g., discussion in Chapter 2 of Bane and Ellwood (1994), supra note 98.
187 For both accounting- and counterfactual simulation-based discussions of the important role that changes in entry rates appear to have played in changing welfare participation, see Jeffrey Grogger, Steven J Haider, and Jacob Klerman (2003), “Why Did the Welfare Rolls Fall During the 1990’s? The Importance of Entry,” American Economic Review, v93, n2 (May), pp. 288-292.
188 Random assignment experiments have many advantages, but the inability of random assignment evaluations to measure entry effects is a serious shortcoming. It has been noted by a number of authors in the evaluation literature.
On the other hand, for women not yet facing a binding time limit, Jobs First actually encourages entry via the generous earnings disregard (see section B above). So it is possible that in the four-year period I study, accounting for entry effects would actually lead me to conclude that Connecticut’s welfare reform caused even of the caseload drop than Table 5 suggests.

The entry-effect problem is likely to be less pronounced in measuring the effects of welfare reform on employment than on welfare participation. Thus, I will repeat the above analysis using employment rates as the outcome variable. Figure 11 plots the time series of employment rates, measured as the fraction of women with positive quarterly earnings, by program assignment group. It shows that employment increased more, and more quickly, among Jobs First group members than among those assigned to AFDC, though AFDC group members also increased their employment over the study period.

Figure 11: Post-Random Assignment Employment, By Jobs First Program Assignment
In Table 6, I quantify these trends. The table’s top row of shows that 39.1 percent of Jobs First group members had positive quarterly earnings at baseline. This share increased by 21.0 percentage points over the 16 quarters following random assignment. AFDC group members had a baseline employment rate of 39.6 percent, which increased by 13.9 percentage points. The difference in these changes is 7.1 percentage points (21.0 minus 13.9). This difference represents only 34 percent of the 21.0 percentage-point increase in the post-random assignment employment rate for Jobs First group members.

Table 6: Key Results on Employment from Jobs First Evaluation

<table>
<thead>
<tr>
<th></th>
<th>Jobs First Group</th>
<th>AFDC Group</th>
<th>Difference in Differences</th>
<th>Share Estimate</th>
</tr>
</thead>
<tbody>
<tr>
<td>Quarter of RA</td>
<td>Change</td>
<td>Quarter of RA</td>
<td>Change</td>
<td>Differences</td>
</tr>
<tr>
<td>Full sample</td>
<td>39.1</td>
<td>39.6</td>
<td>13.9*</td>
<td>7.1*</td>
</tr>
<tr>
<td>Recipients only</td>
<td>35.8</td>
<td>34.6</td>
<td>18.3*</td>
<td>7.0*</td>
</tr>
</tbody>
</table>

*Differences are statistically significant with p values equal to 0 to three digits. “Share Estimate” is the ratio of “Difference in Difference” estimate to Jobs First Group “Change” column.

The table’s bottom row shows that employment results were broadly similar for the subset of women who were receiving AFDC at the time of random assignment. Of these women, 35.8 percent of Jobs First group members had positive quarterly earnings at the time of random assignment, and this share increased by 25.3 percentage points over the 16 quarters following random assignment. AFDC group members had a baseline employment rate of 34.6 percent, which increased by 18.3 percentage points. The difference in these changes is 7.0 percentage points (25.3 minus 18.3). This difference is roughly the same as the difference for the full sample, and it represents only 28 percent of the 25.3 percentage-point increase in the post-random assignment employment rate for Jobs First group members who were receiving AFDC at the time of random assignment.

In the top two rows of Table 7, I repeat the same basic exercise for average quarterly earnings (measured in nominal dollars). In the quarter of random assignment, women in the Jobs First group had average earnings of $586, while those in the AFDC group had average earnings of $631. By the 16th quarter following random assignment, quarterly earnings among those assigned to the Jobs First program had risen dramatically, by $1,651. But the AFDC group saw nearly the same increase, with the $89 difference between these two increases amounting to just five percent of the increase in the Jobs First group’s earnings. The same calculation shows that, of the equally impressive increase in quarterly earnings among incumbent recipients assigned to the Jobs First group, only six percent can be causally attributed to welfare reform. These results indicate that whatever factors caused the Jobs First group’s earnings to increase so much had little to do with changes in welfare rules.

Table 7: Results on Average Quarterly Earnings, Transfers, and Earnings plus Transfers (in nominal dollars)

---

189 I use the time of random assignment as baseline for both the full sample and the recipients subsample. This is appropriate because earnings data do not have the problem discussed in footnote 184.
The next two rows of Table 7 analyze changes in transfer payments, i.e., combined cash assistance and food stamps payments. In the random assignment quarter, Jobs First group members had average transfer payments of $1,579, compared to $1,433 for those in the AFDC group. For Jobs First group members, transfer payments dropped by $1,047 by the 16th quarter following random assignment, which was $235 more than the $812 decline for AFDC group members. Thus, 22 percent (100 percent times 235 ÷ 1,047) of the decline in transfer payments among Jobs First group members was caused by differences in welfare program rules. For incumbent recipients assigned to the Jobs First group, transfer payments fell by $237 between the quarter of random assignment and the 16th quarter following. This amounted to 16 percent of the $1,426 drop observed among Jobs First group members.

Next, consider the bottom two rows of Table 7, which compare changes in the sum of earnings and transfers—what I will call total measured income. The figures in these rows show that Jobs First group members saw their quarterly total measured income rise by $603 between random assignment and the 16th quarter following. This is a substantial increase given their baseline total measured income of $2,165. However, total measured income rose by $750 for AFDC group members over the same period. Thus, the Jobs First program caused quarterly total measured income to fall by $147 over the follow-up period—which means that quarterly total measured income would have been 24 percent greater among Jobs First group members had they instead been assigned to the AFDC group. The same analysis shows that for incumbent recipients assigned to Jobs First, quarterly total measured income among these women rose by only $189 during the follow-up period. The difference-in-differences estimate in the final row of

---

1 Differences not statistically significant from zero at conventional significance levels.
* Differences are statistically significant with p values equal to 0 to three digits.
** Difference is statistically significant with p value equal to 0.068.
*** Difference is statistically significant with p value equal to 0.003.

Note that this measure of income excludes the value of the EITC, unreported earnings, transfers from family members or significant others (including child support), and so on. However, it is the only measure for which the Jobs First evaluation provides directly observed data.
Table 7 shows that these women’s increase in total measured income would have been $134—or 71 percent—greater still had they been assigned to AFDC.

At least for Connecticut, the results in Table 5–Table 7 constitute a striking rebuke to the triumphalist consensus, which holds that welfare reform is the. These results show that reforms in the welfare rules were a relatively small factor in explaining either the drop in caseloads or the increase in employment observed over the period between 1996 and 2000. The results also show that differences in program rules explain essentially none of the very substantial growth in earnings among women in the Jobs First evaluation. If these facts hold for other states as well, that would sharply contradict the view that changes in welfare program rules “were one of the primary causes behind falling caseloads, rising employment, and growing earnings among single mothers,” as Rebecca Blank recently put it. Changes in Connecticut’s program rules may have been a factor, but they have certainly not been a “primary” one.

Moreover, the part of the caseload drop that is attributable to reform appears to be due entirely to the enforcement of Connecticut’s time limit. Not even the most fervent devotees of evaluation methodology needed a random assignment experiment to know that welfare caseloads could be reduced by simply making many would-be participants ineligible.

In addition, it is notable that most of the Jobs First-induced increase in employment occurred in the early quarters following random assignment. Indeed, employment rates in the sixth quarter after random assignment were 54.5% for the Jobs First group and 47.1% for the AFDC group, which amounts to a 7.4 percentage-point impact of the Jobs First program for that quarter. From this quarter until the end of the follow-up period, the two groups’ employment rates increased by virtually identical amounts. Thus, Jobs First’s entire treatment effect on employment had already occurred by the sixth quarter post-random assignment. Recall that no one could lose benefits as a result of the time limit before the end of the seventh quarter following random assignment. Thus, the quarters when Jobs First’s impact on employment unfolded were also those when Jobs First’s earnings disregard carrot would have been most effective. This finding suggests that much of the relatively small salutary impact of welfare reform on employment—as distinct from the upward trend that would have happened anyway—may be due to carrots rather than program-eliminating sticks.

D. Did Welfare Reform “Work”? Final Thoughts Concerning Jobs First

The results just presented concern only welfare applicants or recipients from two offices in a single state. So a skeptic might accept the results above but refuse to believe they apply more generally. But there are several reasons not to be such a skeptic. First, Table 4 suggests that on a number of relevant demographic indicators, the Jobs First experiment’s sample was not too different from the nationwide caseload. Second, in my working paper Much Ado About Relatively Little, I carried out the same type of analysis using data from experimental evaluations.
of reform programs not only in Connecticut, but also in Florida, Iowa, Minnesota, and Vermont.\footnote{For details of the other states’ reform programs’, which vary substantially, I refer interested readers to Table 1 of my working paper, Gelbach (2004), supra note 161, and the discussion, at 8-9, surrounding it.} Most of the estimates suggested that welfare reform explained relatively little of the caseload drop or employment increase observed among experimental treatment group members in these states. The primary exception was for Minnesota’s MFIP program, for which I evaluated two reform programs. One of these had a generous earnings disregard reform, and one had the disregard as well as other changes, most notably including work requirements; neither program included a time limit. For these programs, the shares of employment changes attributable to reform approach or meet the 50 percent mark.

Thus, while Connecticut’s experience with Jobs First surely was not identical to other states’ experiences with their own pre-PRWORA evaluations, I believe the burden should shift. Triumphalists should be expected to justify their position with evidence other than post-welfare-reform-ergo-propter-welfare-reform. At least according to random assignment evaluation data, different programs in different states likely made very different contributions to observed changes in caseloads and employment. And in at least some of these, if not most, the contributions likely were considerably smaller than many seem to think. Finally, over the 48-month follow-up period, Jobs First caused a statistically significant drop in total measured income due to reductions in welfare payments that exceeded the Jobs First-caused increase in earnings.

So what does Connecticut’s experience with Jobs First tell us about whether welfare reform worked? Even for welfare reform’s cheerleaders, the answer is surprisingly far from an obvious yes. The evidence just discussed shows that trends in key outcome variables were more similar than not among AFDC-assigned and Jobs First-assigned women. Moreover, the reduction in total measured income appears to have been concentrated in the bottom part of the total measured income distribution.\footnote{See discussion at 1003-1004, and Figure 8, of Marianne P. Bitler, Jonah B. Gelbach, Hilary W. Hoynes (2006), “What Mean Impacts Miss: Distributional Effects of Welfare Reform Experiments,” American Economic Review, Vol. 96, No. 4 (Sep., 2006), pp. 988-1012.} Presumably this feature will have worsened over time, as the mechanical effects of time limits eliminate eligibility for welfare benefits over time. At least for Connecticut, experimental evidence suggests that radical welfare reform was not the triumph that some believe it to be.

V. Summary

In this chapter, I have discussed the basic elements of the AFDC program created by the Social Security Act of 1935, as well as the many differences between it and the state TANF programs that replaced it following PRWORA. I have discussed the basic economics of AFDC’s incentive effects, as well as the ways in which TANF programs might be expected to alter those incentives, and thus behavior. I have argued that the empirical evidence on AFDC’s behavioral impact was never nearly as compelling, nor as sizable, as AFDC’s critics maintained. Likewise, there is—at most—limited empirical evidence supporting what I have called the “triumphalist” position, that the wave of state and federal welfare reforms that took place in the 1990s play an
important role in explaining the increase in employment among single mothers. Either the economic boom of the late 1990s, or some other difficult-to-measure change, appears to have played a more important role than welfare reform. What is more, the uncontrolled experimentation of the last two decades has left us with a crazyquilt collection of state programs that do not lend themselves to the kind of empirical research that would be necessary to benefit from the Brandeisian idea of states as laboratories of democracy.