10-1-2005

Welfare Caseloads and the 2001 Recession

Andrew Goodman-Bacon
Macalester College, agoodmanbacon@gmail.com

WELFARE CASELOADS AND THE 2001 RECESSION[*]

Andrew Goodman-Bacon

Macalester College
Economics Department

Karine Moe, adviser

Submitted May 2, 2005

Abstract: *This paper investigates the effects of welfare reform policies on the number of families receiving welfare (caseloads) since the 2001 recession. 1996's welfare reform legislation was passed amidst the longest economic expansion in US history, making it hard for researchers to estimate the role of policy. Furthermore, caseload research has not sufficiently explored the effects of specific policy choices within a broader reform package. This paper uses state panel data to examine the effects of specific policies on caseloads since the recession. Results indicate that since the 2001 recession, generous financial incentives to work reduced the number of families on welfare, while time limits and punishments for non-compliance had no impact. Taken as a whole, welfare reform helped move low-income women off of welfare in the 90s boom and subsequent recession.*

[*] Many thanks to Karine Moe, Gary Krueger, Sarah West, Peter Ferderer, Matija Vodopivec, and the Macalester Economics Department. I thank Bob Schoeni for his generous guidance, and John Horejsi and Evelyn Mills of HHS for providing caseload data.

Table of Contents

I. Introduction

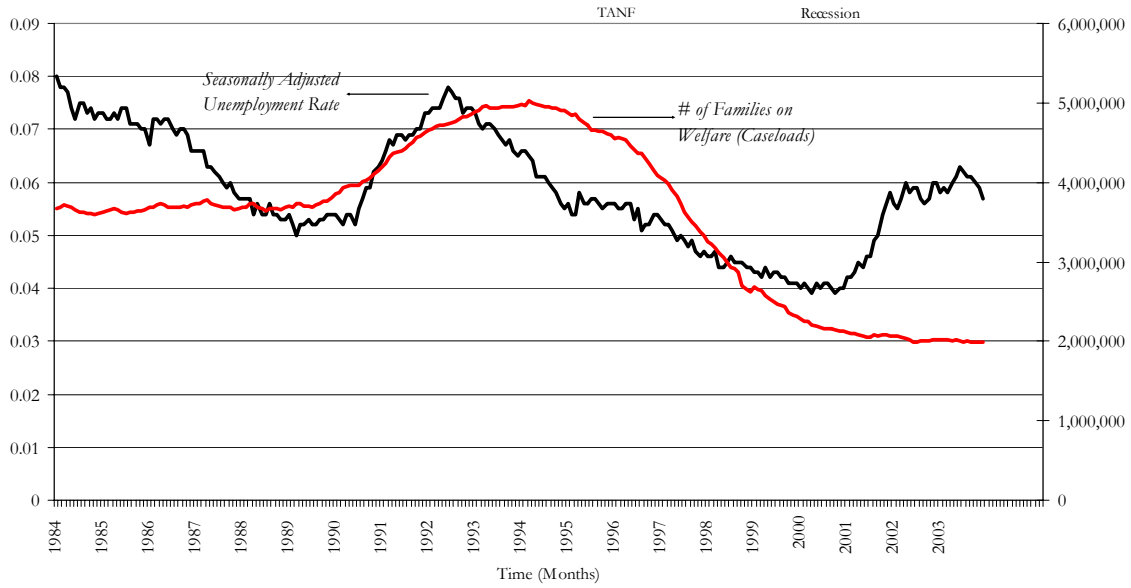
Since the mid-1990s families have moved off welfare at an unprecedented rate, leaving the level of program participation at 39% of its historical peak. This phenomenon represents a dramatic shift in the behavior of low-income single mothers. Economists attribute this caseload decline to welfare reform and the strong macroeconomy, but fail to conclusively identify the relative contribution of these factors or of specific policies within the reform. Such identification is difficult because both welfare reform and economic conditions affected caseloads in the same direction over the 90s. The recession in 2001 provides the necessary macroeconomic variation to address this shortcoming and improve policy analyses.

Caseload reduction was an explicit goal of 1996's Personal Responsibility and Work Opportunity Reconciliation Act (PRWORA), and it remains a common measure of welfare reform's success. While caseloads usually move with the business cycle, they responded mildly to the 2001 recession, the first of the post-reform era, which suggests that policy played a role. Figure 1 plots national caseload levels and the unemployment rate from 1984 to 2003. About 110,000 fewer families received welfare in December 2003 than at unemployment's trough[1] —a very small change relative to the 1990s, when 3 million families left welfare, but certainly not an increase. Caseloads did not continue their decline in all states, though, and since welfare programs vary by state, the recession is a useful natural experiment for analyzing policy.

[1] The National Bureau of Economic Research calls March the first month of the 2001 recession. http://www.nber.org/cycles.html/

Figure 1

Families on Welfare and the Unemployment Rate (1984-2003)



Source: US Department of Health and Human Services and US Bureau of Labor Statistics[2]

This paper examines the effects of welfare reform on caseload changes since the 2001 recession. This should give better estimates of welfare reform's role in a long-run trend in caseloads and in short-run fluctuations around that trend. Results indicate that in boom and bust periods, welfare reform reduces caseloads but that certain aspects of policy affect them quite differently. Section II presents the theoretical model, Section III reviews the literature, Section IV presents the empirical framework, Section V presents results and Section VI concludes.

II. Theory

Since the New Deal, welfare has been a federal transfer program for low-income single mothers. Before 1996, as long as women met these eligibility requirements they were entitled to a monthly benefit check. This unlimited entitlement led to dissatisfaction

[2] 3 observations were missing for Ohio so all caseload graphs show a slight dip in 1998. These observations were dropped when estimating the regressions, so they do not alter these results.

with recipients' low levels of work and perceived dependence on welfare in the 70s (Moffit, 1992), and calls for welfare reform (Levy, 1979). Beginning mainly in the early 90s, the federal government granted waivers to states to experiment with small scale reforms. These experiments led to Temporary Assistance for Needy Families (TANF), the federal block grant program established under PRWORA to replace Aid to Families with Dependent Children (AFDC). TANF ended entitlement to federal welfare benefits and gave states greater freedom to design their own programs, subject to general federal guidelines:[3]

- A 60-month lifetime limit on benefits
- Work requirements for annually increasing percentages of states' caseloads.

Time limits sought to end welfare dependence, and work requirements sent the message that employment was the path out of poverty.[4]

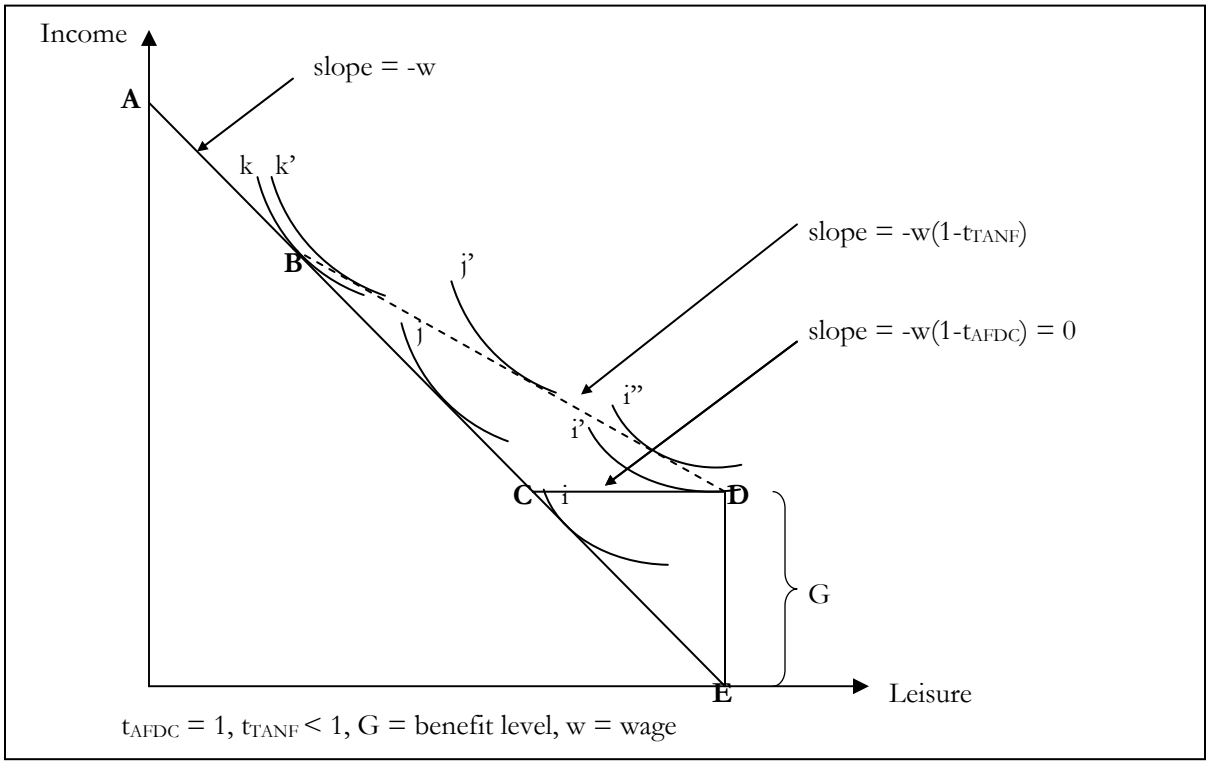
States' TANF policies mainly differ by four basic program parameters:

- *Benefit Levels:* the size of the cash grant given to welfare recipients, denoted by 'G'[5]
- *Earnings Disregards:* the rate at which participants retain their benefits as they earn more, denoted by 't'[6]
- *Time Limits:* the number of months women may receive welfare over their life time,
- *Sanctions:* the benefit reduction for non-compliance with program rules, usually work requirements.

[3] For more information on federal and state welfare policy see an overview by Gil Crouse at the Department of Health and Human Services website: http://aspe.hhs.gov/hsp/Waiver-Policies99/policy_CEA.htm.
[4] It is important to note that apart from the strong macroeconomy, expansion of the Earned Income Tax Credit (EITC) and increases in the minimum wage also drew low-income people into the labor market, further complicating an empirical identification of welfare policy effects.
[5] Benefits range from around $120 per month in Alabama to more than $900 per month in Alaska. While states were free to choose their benefit levels before TANF, there is evidence that variation in this parameter increased. After TANF's passage, the standard deviation of benefit levels increases from $151 to $159.
[6] Benefit retention has often been expressed by the equivalent concept of an implicit tax on earnings, t, also called the benefit reduction rate. Because benefits are calculated continuously (see fn. 3), "disregarding" a portion of earnings in that calculation means that as recipients earn, they keep some benefits. This is equivalent to saying that as recipients earn, their income (earnings plus benefits) are taxed at some rate until the benefit equals 0. If the earnings disregard increases (decreases), t falls (rises). Both mean that a larger (smaller) portion of G is retained as recipients work.

The standard labor-leisure model shows how these program parameters change the work incentives welfare recipients face. Figure 2 illustrates the predicted effects of changes in benefit levels and earnings disregards. Both axes represent time in hours, but the income axis measures time*wage to show income as a function of hours worked. It is assumed that there are no other taxes or transfers programs.

Figure 2



Indifference curves *i*, *j* and *k* represent three single mothers. AE is the budget constraint in the absence of any program and its slope is the wage, *w*. Welfare provides a cash benefit, *G*, when income and hours worked are zero (point E). A key aspect of transfer programs, though, is the rate, *t*, at which that benefit falls as recipients' earnings increase. This rate has been of concern to economists at least as far back as Milton Friedman's *Capitalism and Freedom* (1962) and was examined extensively in the 70's and 80's (Levy, 1979; Ashenfelter, 1978 and 1983; Moffit, 1986; Leonisio, 1988; Moffit,

1992). *t* affects the marginal benefit of work for welfare recipients by determining the effect on total income of an additional hour worked. For recipients, income increases by *w(1-t)* for each hour worked, so both the wage and the tax rate matter to total income. Second, this rate changes the break-even level of income at which recipients become ineligible. The break-even level is *G/t*.[7] Therefore, when *t* falls, the break-even level rises and more women are eligible for welfare.

According to Ashenfelter (1978), when policy changes eligibility, caseloads will change "mechanically." New eligibles participate simply because they are eligible without having to change their work behavior to do so. When caseloads change because work behavior changes, these are called "behavioral" effects. All the program parameters listed above have both mechanical and behavioral effects, which I discuss below.

Under AFDC, benefits were reduced one-for-one with earnings (t_{AFDC} was equal to one). This is shown in Figure 2 by budget constraint ABCDE.[8] Notice by indifference curves *i* and *i'* that this reductions leads recipients to not work at all and to receive the benefit as a windfall. When *t* is one, the marginal benefit of work is zero (*w(1-1) = 0*) because each dollar earned is offset by a dollar lost in benefits. Setting *t* to one also sets the break-even level of income to its lowest possible level (assuming that *t* cannot be greater than one), *G*. Thus, after earning relatively little, recipients lose eligibility and leave the rolls.

---

[7] Let H be hours worked and y* be the net transfer. The net transfer is calculated by:

$$y^* = G - (wH)t$$

If wH = G/t, we see that y* = 0. Therefore, G/t is the break-even level of earnings.

[8] From 1967 to 1981 the implicit tax rate, *t*, fell to 67%, but 1981's Omnibus Budget Reconciliation Act restored it to 100% (set earnings disregards to 0). From then on, "without a waiver, individuals [were] allowed to keep $30 plus one-third of all additional earnings for the first three months of benefit receipt (the 'standard AFDC disregard'). After that almost every dollar of earnings result[ed] in a dollar reduction in benefits." CEA, 1997 p.16

Under TANF, states increased earnings disregards (i.e., reduced *t*), allowing recipients with higher earnings to receive benefits. This is shown in Figure 2 by the dashed portion of the TANF budget line, ABDE. As under AFDC, the slope of the dashed portion is $-w(1-t)$, but because t_{TANF} was less than one the marginal benefit of work is greater than zero. Also, the break-even level of income increases when *t* falls so more women are made eligible. In terms of participation and work effort, increased earnings disregards affect several groups (Blank, Card and Robins, 2000). Furthermore, the short and long-run caseload effects of earnings disregards may differ for some groups.

First, women who received AFDC and did not work (indifference curve i') will continue to receive welfare, but will likely work more. TANF's earnings disregards increased their effective wage. The substitution effect leads to more work hours, but the income effect cannot lead them to work less because they already work zero hours. An increase in work effort is shown by indifference curves *i'* and *i''*. Notice that work effort with higher earnings disregards is still lower than in the absence of welfare. In the short-run, then, this parameter should not affect caseloads for non-working AFDC recipients, but if the increase in work effort leads to steady employment or wage advancement then higher earnings disregards may ultimately reduce welfare participation among these women.[9] This group made up a large portion of AFDC recipients. Between 1968 and 1987 no more than 18% and as few as 5% of recipients worked at a given time (Moffit, 1992). Therefore, long-run effects may be substantial.

Second, some women who were not eligible for AFDC but are eligible for TANF will choose to participate, but they will reduce work hours. These women originally locate between C, the AFDC break-even level, and B, the TANF break-even level

[9] Gladden and Taber (2001) present evidence that low-skill workers experience wage advancement with experience.

(indifference curve *j*). The cash grant reduces work hours by the income effect and a reduction in the effective wage (to *w(1-t)*) reduces work hours by the substitution effect (indifference curve *j'*). Higher earnings disregards will increase caseloads mechanically among these "windfall" beneficiaries in the short run. It is unclear whether or not their reduction in work hours will affect a long-run scenario in which work experience leads them off welfare by a behavioral effect.

Third, an increase in the break-even level may lead some women, locating just above B, to reduce work hours and receive welfare. A reduction in hours in order to use welfare is shown in Figure 2 by indifference curves *k* and *k'*. Like the "windfall" beneficiaries, caseloads should increase in the short-run among these women, but a long-run effect is ambiguous.

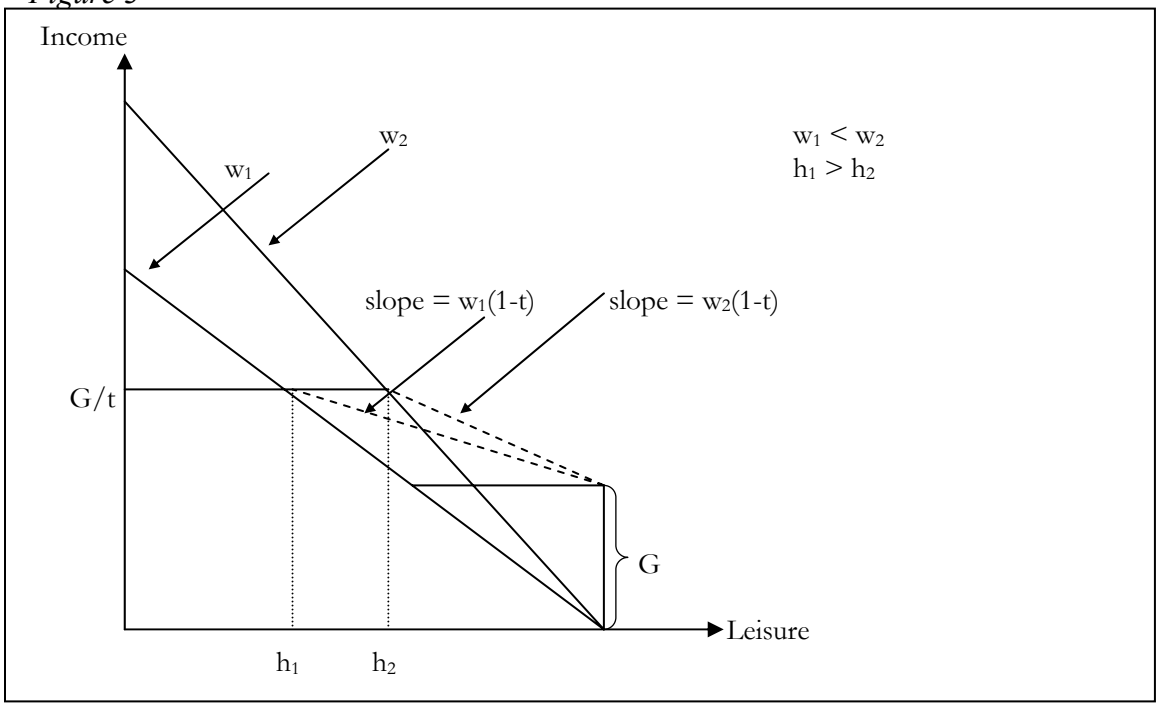
In Figure 2, time limits and sanctions are exogenous incentives for recipients to leave welfare. Strictly speaking, sanctions, generally levied on recipients who do not meet TANF's work requirements, encourage work. If a participant is sanctioned, they lose some portion of their benefits, and the dashed TANF budget line will shift down. Therefore, sanctions have two caseload reducing effects. First, since participants want to avoid a backward shift in their budget constraint, they should tend to follow the rules and work—the behavioral effect. Insofar as working leads people off welfare, sanctions should reduce caseloads. Second, if sanctions are levied on non-compliant recipients, caseloads will fall when *G*, and thus the break-even level, *G/t*, falls—the mechanical effect.

Time limits, on the other hand, only address welfare receipt. If a participant reaches the 60 month time limit, her benefits will be cut off and her budget line will

become AE. Time limited recipients have an incentive to leave the program, whether or not they are working, in order to save eligible months—the behavioral effect. When recipients reach a time limit, caseloads fall—the mechanical effect. Overall, time limits should reduce caseloads.

The strong macroeconomy of the 1990's also changed work incentives and decisions about welfare use by making employment plentiful and rewarding. The national unemployment rate was at or below 5% from May, 1997 through September, 2001 making jobs easier to find for low skill women. In fact, over the 1990's labor force participation rates increased 5% among female high school dropouts, 10% among single mothers, and 25% among women who received public assistance. Low skill wages began to rise in 1995, and this trend was particularly strong among single mothers (Blank and Schmidt, 2001). Theoretically, higher wages should reduce caseloads because the number of hours required to earn the break-even level of income falls. Figure 3 shows the effect of rising wages on the break-even number of hours.

Figure 3



When the wage rises from w_1 to w_2, G/t is unaffected, but it is clear that the hours needed to earn G/t falls from h_1 to h_2. Thus, caseloads should fall when wages rise. This should also increase work effort among those eligible for welfare because the substitution effect tends to dominate at low wages.

Clearly, both economic conditions and welfare policy led low income women to work rather than receive welfare over the 1990's. Therefore, a robust analysis of TANF's effect on program participation must consider times of recession as well as prosperity.

III. Literature Review

Caseload research has addressed two main questions: what primarily drives caseload change? And what were the relative contributions of the strong economy and welfare policy to caseload changes in the 1990's? Here I focus mainly on the methodological aspects of the first question. The majority of caseload studies (CEA, 1997; Wallace and Blank, 1999; Bartik and Eberts, 1999; Moffit, 1999; CEA, 1999; Figlio and Ziliak, 1999; Blank, 2000; Schoeni and Blank, 2000; Grogger, 2000; Danielson and Klerman, 2004) use state-level administrative panel data to examine caseload levels.[10] Three studies, Wallace and Blank (1999), Ziliak et. al (2000) and Danielson and Klerman (2004) use monthly data. The first two present evidence of nonstationarity in monthly caseloads and estimate first-differenced models, while Danielson and Klerman reject the null hypothesis of a unit root and correct for autocorrelation with an AR(12) model. The most basic model (CEA, 1997) considers

[10] Moffit (1999) and Schoeni and Blank (2000) are exceptions. They use individual level CPS data rather than statewide administrative data.

caseloads to be a function of economic conditions and welfare policy, using state and time fixed effects to capture unmeasured factors.

III.a. Measuring Economic Conditions

The unemployment rate produces strikingly robust effects as a measure of the economic conditions faced by welfare recipients. A 1% increase in the unemployment rate tends to increase caseloads between 4 and 7 % over two years. Two studies examine the degree to which the unemployment rate directly affects the functions of reform. That is, if welfare reform encouraged recipients to enter the labor market, then caseloads should have become more cyclically sensitive under TANF. Moffit (1999) and CEA (1999) regress caseloads on unemployment (among other controls) in separate time periods and find that caseloads were more sensitive to unemployment over the 90's than in other time periods. These results warrant further analysis, though, because caseloads did not respond to the jump in unemployment in 2001. The unemployment rate alone, while an obvious choice for measuring labor market conditions, may not fully capture the labor market incentives welfare recipients face.

Bartik and Eberts (1999) criticize the unemployment rate and other "variables that simply measure the economic status of welfare recipients," as endogenous to welfare policies. They prefer to approximate the labor market conditions faced by welfare recipients rather than measure their reaction to those conditions (working or not working). Bartik and Eberts find that positive state employment growth reduces caseloads, but that a higher fraction of a state's employees with a high school diploma (measuring the educational requirements of that state's industry) and a higher propensity

of a state's industries to hire welfare recipients increase caseloads.[11] Including these variables reduces the unemployment rate's effect. Nevertheless, the unemployment rate remains the standard measure of economic conditions and may accurately reflect economic realities for welfare eligibles given their sensitivity to aggregate macro-measures (Hoynes, 2000).

III.b. Measuring Welfare Policy

To measure welfare reform, studies either include 'reform' dummy variables for TANF and any major waiver, or several dummy variables for the specific policy parameters discussed in section II.[12] 'Reform' dummies measure the extent to which the momentum of welfare reform (policies *and* attitudes about ending entitlement and stressing work) reduced caseloads, but bury the effects of policy parameters within that rather broad context. 'Policy' dummies attempt to isolate the individual impact of program parameters, but indicate less clearly whether the economy or reform primarily drives caseload changes.

Neither measurement method precisely estimates waiver effects. 'Any major waiver' dummies are generally found to reduce caseloads, but not always. They account for no more than a third of the caseload decline between 1994 and 1996, while they are able to explain little to none of the caseload increase before 1994.[13] CEA (1997) and Moffit (1999) characterize a state's waiver as only one of six policies: sanctions, work exemptions (for women with young children), time limits, earnings disregards, benefit

[11] The authors attribute the negative effect of a state's willingness to hire welfare recipients to high turnover rates in those states' industries, guessing that while they may hire welfare recipients at a high rate, they also shed employees at a high rate, creating even more welfare recipients.
[12] Benefit levels are not included as dummies, but as the natural logarithm of their real state-year value.
[13] See Figure 1. Moffit (1999) finds that waivers increase caseloads as the relationship between the unemployment rate and caseloads is more accurately estimated. Blank (2000) finds similar results when she adds state time trends. This sensitivity is likely a result of diversity among waivers' expected effects.

levels and family caps (limits on benefits paid to recipients who have children while on the rolls). Their results are not robust in that no policy variable is significant in both studies except benefit level. This result is unsurprising because these models characterize a waiver as one policy type while states actually implemented a combination of caseload reducing (stricter sanctions and time limits) and caseload increasing (higher earnings disregards) waivers.

Unambiguously, studies using a single reform dummy find that TANF has a larger negative effect on caseloads than do waivers. Estimates of this effect range from -12 to -50 percentage points. Unfortunately, aside from Danielson and Klerman (2004) and CEA (1999), specific policy choices under TANF have not been examined.

Time limits are found to reduce caseloads, although their effect is rarely significant because few studies include observations after time limits became binding. Thus they only estimate the behavioral effects discussed in section II. Danielson and Klerman (2004) use data through June 2003, a range that includes the first month when time limits were effective in most states. They find large effects, both behavioral and mechanical, summing to a 22% reduction in caseloads after full-family time limits become binding. A further complication in estimating time limit effects is that not all time limits terminate cases. Some reduce benefits, while others simply require that a recipient begin working. Smaller benefit reductions and longer time frames have smaller effects. Grogger (2000) provides evidence that time limits have greater behavioral effects for families with younger children, because they are potentially eligible for welfare for longer than the time limit (families lose eligibility when the youngest child reaches 18) and should, therefore, conserve months.

Other policy conclusions are less detailed and accord with theory. Higher benefits and higher earnings disregards are found to increase caseloads and the stronger a state's sanction policy the more its caseloads fell. These results do not distinguish between behavioral and mechanical effects.

III.c. Additional Variables

Even with detailed policy characterizations, the CEA model is quite sparse, relying on state and year fixed effects to capture changes in many factors that influence caseloads. Several studies explore omitted variables such as other anti-poverty policies, wages, and demographics.

Wallace and Blank (1999) show that welfare participation is a significant predictor of food stamp rolls, indicating that program interaction is important. The EITC expansion, for instance, had significant caseload reducing effects even after controlling for welfare reform (Grogger, 2003).[14] Interaction between types of welfare caseloads is also important (Blank, 2000). Welfare predominantly serves poor single mothers, but it also provides benefits, on a relatively limited basis, to two parent families and to children but not their legal guardians. She finds that two parent cases follow the business cycle more closely than single-mother cases, while child-only cases are poorly explained by her model. Accounting for them separately, though, greatly improves its explanatory power.

Figure 2 shows that earnings affect program participation via earnings disregards. Blank (2000) and CEA (1999) find that Median, 20th percentile, 10th percentile and state minimum wages all have significant negative effects on caseloads. In fact, increases in

[14] Grogger (2003) is part of a second branch of caseload literature that uses individual level longitudinal data to estimate welfare entry and exit rates (caseload flows) with logit models. Since this branch of research is small relative to "stock studies" (those discussed above) and my data do not permit the estimation of flow models, I omit discussion of their methodology. It is worth noting that this method does not significantly or uniformly change any conclusions about the determinants of caseload change.

20th and 10th percentile wages add to the downward effect of increases in the median, indicating that wage distribution is important to caseload levels. That is, shifts in the entire income distribution matter, but so does the shape of the low end of the distribution.

Blank (2000)[15] includes an expanded set of demographic controls in an analysis of AFDC caseloads. She finds statistically significant results for: states' share of non-marital births, single female heads and immigrants, which increase caseloads; and average years of education and elderly share, which reduce caseloads. The share of a state's population that is black increases caseloads, but this result is not precisely estimated. Republican governors and politically unified state legislatures reduce caseloads. Including these demographic and policy variables improves the specification, but, importantly, does not change the conclusions of sparser models.

III.d. Relative Contributions of Macroeconomy and TANF

Recent caseload research also investigates the relative contributions of the strong economy and welfare policy to caseload changes in the 1990's. Economic conditions appear to have mattered less to caseload declines under TANF, although this is not to say that, at the margin, cases are less sensitive to these conditions. Rather, this conclusion is a function of the rates of change of policy and unemployment over the 90's. PRWORA's reforms were much broader than waiver-based reforms and, all things equal, should have had a larger effect relative to the economy. Also, after 1996, unemployment was falling less quickly than in the early 90's so it played a smaller role in caseload changes.

[15] Rebecca Blank's contribution to the welfare reform literature warrants a special note here. She is involved in most of the best work in this area, and this paper relies heavily on several of her conclusions and methodologies.

While the caseload literature is large relative to other welfare research, it does a poor job of exploring specific effects and has not widely considered the impact of the recession. This paper addresses both of these issues.

IV. Empirical Strategy:

IV.a. Econometric Specification

I consider caseloads to be a function of economic conditions, policy parameters and demographics. Figures 2 and 3 show how economic conditions and policy parameters affect individual decisions about welfare use and work. Demographics affect aggregate caseloads to the extent that they determine the number and nature of individual decision makers. For instance, if African-Americans and women face wage gaps in the US, then a state with more of black women should have more people who tend to have lower wages, more eligibles and, therefore, higher caseloads. Population is the simplest example: states with higher populations have more welfare recipients.

I take much of my methodology from Wallace and Blank's (1999) differenced model (their equation 2): [16]

$$\Delta \ln(R_{st}) = \Sigma \beta_j * \Delta U_{st,j-q} + \Sigma \alpha_j * \Delta P_{st,\, j-q} + \Sigma \Delta sm_{st} + \Delta \varepsilon_{st} \qquad \text{(eq.2)}$$

where R is the number of families on welfare divided by the population, U is the unemployment rate, P is a vector of reform dummies, s is the state, t is the month, sm_{st} are state-month effects (51*12 = 612 dummies) and ε_{st} is a stochastic error term. 'q' is the lag length, set to 11 and 23 months, and all summations index j from 0 to q. State fixed effects are linear combinations of state-month effects so this model corrects for

[16] At a monthly frequency non-stationarity in caseloads is evident. Augmented Dickey-Fuller tests cannot reject the null hypothesis of a unit root for most states, and at any given time around 80% of welfare recipients live in states where the caseload is non-stationary. Therefore, first differences are appropriate.

strictly cross-sectional variation. State-month effects also control for seasonality in caseloads.

It should be noted that state-month effects are not also equivalent to time fixed effects, which would control for omitted federal policy variables such as the EITC and minimum wage changes that took place over the 1990s. Since Wallace and Blank's model does not control for this variation, their policy effects may be biased upward if other policy changes drew women off welfare. I maintain this specification, and therefore this potential bias, because state-month effects are collinear with time fixed effects. As such, my policy conclusions may not be attributable purely to TANF and should be interpreted as TANF's effect *conditional on* other policy changes.[17] Note that unless the omitted variables are correlated with welfare policies across states, the relationship between specific policy coefficients is still meaningful.

First differencing poses a challenge for policy dummy variables. Differenced dummies equal zero in all months except the month in which they turn on, so they only affect the dependent variable in that month. TANF affects caseloads in every month since its implementation, though, so differencing a single TANF dummy misses most of its impact. Wallace and Blank's solution is to lag the differenced TANF dummy until the end of their time period. Each coefficient, therefore, gives the estimated change in caseloads associated with TANF for *that* month. This policy specification allows its effect to differ over time. This flexibility is desirable, but as the date range grows the number of lags must also grow. Not doing so underestimates the change in caseloads

[17] Blank (2002) notes that the biggest EITC expansion and minimum wage increases took place *before* PRWORA was passed, so all TANF policy coefficients should, theoretically, suffer from the same bias. Therefore, comparing the signs and magnitudes of my policy coefficients is appropriate if we assume the omitted variable bias is the same across states.

after the lag length ends.[18] Eventually, adding sufficient TANF lags consumes too many degrees of freedom, especially if multiple policy variables rather than a single TANF dummy must be lagged.

To save degrees of freedom while fully accounting for TANF, I use undifferenced policy dummies that equal one in every month a particular policy ('reform' in general or specific policies) was in place in a state, instead of policy lags.[19] Since there is no lag involved, each variable produces one coefficient. This coefficient represents the change in caseloads that took place in every month that a state had that particular policy, or the change in the caseload *trend* associated with that policy. One disadvantage is that OLS applies this coefficient to every month, when the true policy effect probably differs across months. An advantage is that this structure makes the coefficients on different policy variables directly comparable. Multiplying the coefficient times the total number of months that the policy was in effect gives the total estimated impact of that policy, holding all other factors constant. Thus, the final model is of the form:

$$\Delta \ln(R_{st}) = \Sigma \beta_j * \Delta U_{st,j-q} + \Sigma \delta_j * \Delta W_{st} + \alpha * P_{st} + \Sigma \Delta sm_{st} + \Delta \varepsilon_{st} \qquad \text{(eq.5)}$$

where all variables are defined as in equation 2, but P_{st} is a vector of undifferenced policy dummies. Data limitations prevent the inclusion of demographic variables, but Blank (2000) shows that this may not substantively change any conclusions.

My policy specification builds on previous research by accounting for different levels of specific policy choices and allowing policy effects to differ in boom and bust

[18] In fact, if Wallace and Blank's specification is run with more dates, a shift in the residuals is evident. Figure 1 in the data appendix shows this residual plot, Figure 2 shows one with undifferenced policy dummies. The mean of the residuals after the 23rd month of TANF is -0.001, compared to -0.006 before the 23rd month. This difference changes when the month used to split the residuals changes.

[19] Wallace and Blank note that including an undifferenced dummy "represents a shift in the rate of change in caseloads," (p.73). Hoxby (2001) employs a similar methodology to estimate the impact of charter school competition on public school performance. She differences the dependent variable and regresses it on a set of undifferenced dummies.

periods. The longer TANF is in effect, accounting for diverse policies becomes more important. While a single TANF dummy captures the initial momentum of such a drastic, national policy change, we need additional variables to control for policy variation as that momentum fades. Also, subsequent welfare policy decisions will deal with specific policy choices, not reforms of TANF's magnitude. In addition to a general TANF dummy, I use specific policy dummies for overall work incentives, earnings disregards, time limits and sanctions. The overall work incentives categorization follows Blank and Schmidt (2001), who consider a state's entire set of program choices and label their work incentives as "strong", "mixed" or "weak."[20] Earnings disregards are either "low", "medium" or "high", corresponding to generosity. Sanctions and time limits are either "lenient", "moderate" or "strict" referring to the swiftness of sanctions, the length of time limits, and the amount of benefits taken away from sanctioned/time limited recipients.

Even though policy parameters vary much more under TANF than they did under AFDC, certain policy choices are correlated. Table 1 shows correlation coefficients for policy parameters across states.

Table1 - Correlation Matrix of Policy Parameters

		Earnings Disregards			Sanctions		
		Low	Medium	High	Strict	Moderate	Lenient
Sanctions	Strict	0.28	-0.06	-0.23	1		
	Moderate	0.11	-0.23	0.12	--	1	
	Lenient	-0.38	0.25	0.15	--	--	1
Time Limits	Lenient	-0.04	-0.09	0.13	-0.23	0.09	0.18
	Moderate	0.06	0.10	-0.16	-0.15	0.01	0.10
	Strict	-0.03	-0.03	0.06	0.35	-0.09	-0.25

[20] If a state is "strong" ("weak") in at least one parameter and "weak" ("strong") in none, then that state is considered to have "strong" ("weak") work incentives.

Stricter and less generous policies are rarely combined with more lenient or generous policies. For instance, the correlation coefficient between low earnings disregards (the least generous) and lenient sanctions (the least strict) is -.38, but the coefficient between low earnings disregards and strict sanctions is .28. While policy variation increased after TANF, the shortage of certain parameter combinations means that estimating specific policy effects is still difficult.

This range of policies is, perhaps, the most important concept to capture accurately yet they are not consistently measured. The Blank and Schmidt policy categorizations (and similar policy measures) have not been updated since 2001, which means that my specification neglects changes to welfare within a state over time. This may not be a serious problem, though; according to state TANF reports to the Department of Health and Human Services policy has changed substantially. Between 2002 and 2004 two states changed their sanction time frame (one became stricter, the other more lenient), two states adopted greater sanction benefit reductions, and three (two) states implemented more (less) generous earnings disregards. Constancy in welfare policy is likely given its political sensitivity. For example, Congress has been extremely reluctant to reauthorize TANF—the law has simply been extended every fiscal year since 2001.

IV.b. Data

I use monthly caseload data (June 1989 through December 2003) from the US Department of Health and Human Services and resident state population estimates from the US Census Bureau to calculate the dependent variable. I use families rather than individuals receiving welfare because this more accurately captures household decisions

about program participation. The number of individual recipients tends to confuse behavioral changes (whole families choosing to receive welfare) with fertility changes (existing families having more children). Since welfare eligibility is determined on a monthly basis, this frequency is appropriate for measuring caseload change.

Unemployment rates are seasonally adjusted and published by the US Bureau of Labor Statistics. I calculate state-month 20th percentile weekly wages from the Current Population Survey's Outgoing Rotation Group, using ORG weights to extend these estimates to a state-wide level. Earnings disregards, sanctions, time limits and work incentive dummies are from Blank and Schmidt (2001, pp.84-85). TANF and waiver implementation dates are from CEA (1999). Benefit levels are from the various editions of the House of Representatives Green Book and the Welfare Rules Database. There are 180 observations per state, and 9,177 overall (3 caseload observations missing for Ohio are dropped). Table 2 presents summary statistics.

Table 2 - Summary Statistics

n = 9,177	Mean	Standard Deviation	Standard Deviation Across State Averages
Families Receiving Welfare	69,743	114,287	108,789
Families Receiving Welfare per capita	0.01	0.007	0.005
Unemployment Rate	5.29	1.51	0.98
Wage Benefit Ratio	.65	.30	0.27

IV.c. Predictions

The relationship between the unemployment rate and caseloads is central to an examination of TANF's effect in recession. I expect unemployment to be positively related to caseloads, but I also expect its magnitude to be smaller than in previous research. Caseloads follow the unemployment rate fairly closely over most of this time period (see Figure 1). The deviation from this relationship in the mid-80's can be

partially attributed to legislated eligibility changes. The mild reaction of national caseloads to the 2.5 point increase in unemployment rates from April 2000 to June 2003, though, suggests that caseloads have become less sensitive to unemployment rates.

Several factors may be at work. First, TANF pushes recipients to use the labor market much more than did AFDC. Thus, participation should be more sensitive to labor market fluctuations after reform, and there is evidence to support this claim. Another component of welfare reform, though, was a get-tough attitude with corresponding policies like time limits and sanctions. To the extent that states have followed through on these threats and ended non-compliant cases, some low-income women may not be able to access welfare at all. This would make caseloads less responsive to the unemployment rate. Second, if lifetime limits have forced low-income women to try to preserve available months on welfare, they should be less inclined to receive benefits. This implies a weaker relationship between unemployment and welfare, because time limits make welfare less of a safety net. Third, this recession came 4-5 years after TANF was implemented in most states and 6-7 years after the national caseload started to decline. If recipients worked consistently during this time they may have gained either enough labor market experience or job prestige to quickly find a job or retain employment in the recession, and therefore avoid welfare. In other words, TANF may have worked.

Following the theoretical predictions in section II, I expect stricter time limits and sanctions to reduce caseloads because they provide exogenous incentives for recipients to leave welfare or begin working. I expect higher earnings disregards and benefits to increase caseloads in the short run. If more generous earnings disregards are found to

reduce caseloads, though, I take it as evidence that the long-run effect of the increased work effort of old AFDC recipients dominates the short-run caseload increasing effects.

V. Results

I now present empirical results of policy's effect on caseloads since the 2001 recession. First, I provide graphical evidence indicating that policy parameters explain caseload changes since the recession. Second, I present regression results that build on the monthly specifications in Wallace and Blank (1999). The key results are from the models that allow specific policy choices to have a different effect in the boom and the bust. The findings imply that generous earnings disregards have been effective at reducing caseloads since the 2001 recession, while sanctions and time limits have not. Finally, I compare actual national caseload levels to predicted values using the boom/bust specification with policy parameters.

V.a. Graphical Analysis-

Figure 4

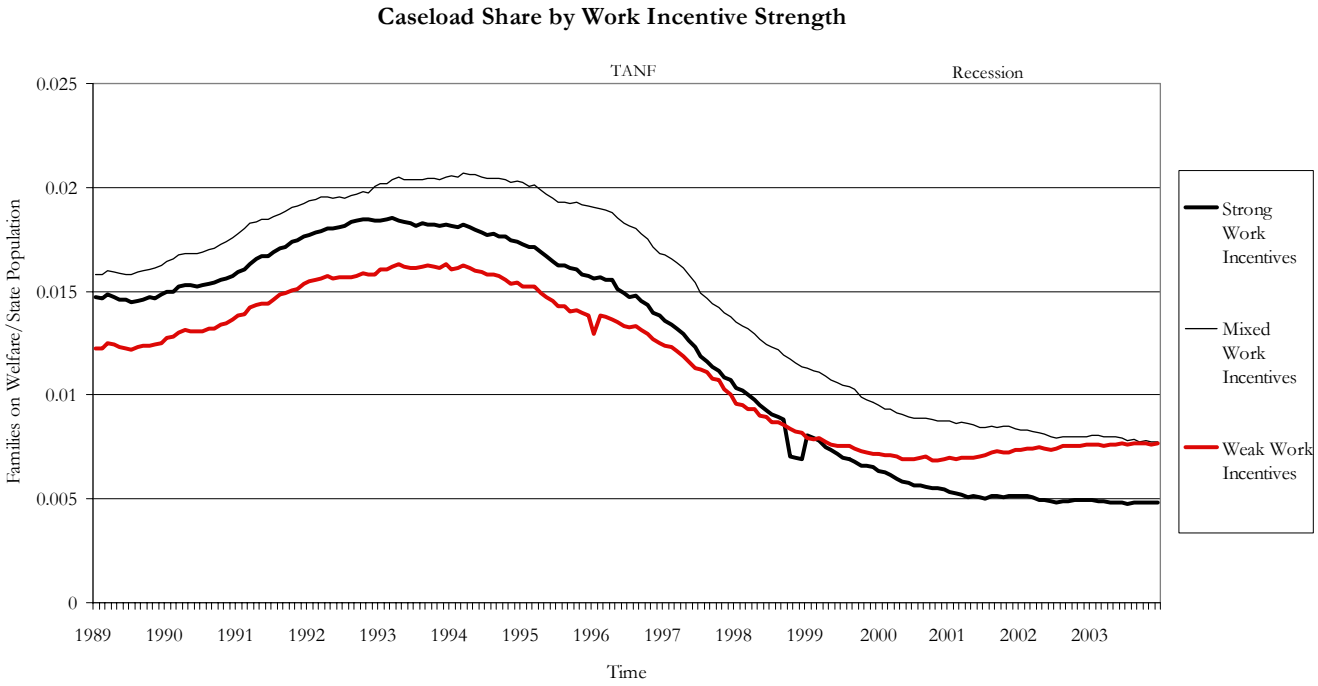


Figure 4 plots caseload shares between 1989 and 2003 by TANF work incentive groups. Each series is the sum of cases in states with each policy, divided by the sum of those states' populations. While all groups experienced the drastic reductions observed nationally, trends by policy are clear. Strong-work-incentive states' ("strong states") caseloads fell at a faster rate after TANF than either of the other two groups, but moved with mixed incentive states in the 00s. By 1999, strong states had a lower caseload share than weak states for the first time. After this point, both strong and mixed states continued to reduce caseloads slightly, while caseloads in weak states rose. This rise is roughly coincident with the recession. Figure 4 provides at least initial evidence that recipients respond to work incentives in recession as well as boom times.

Figure5

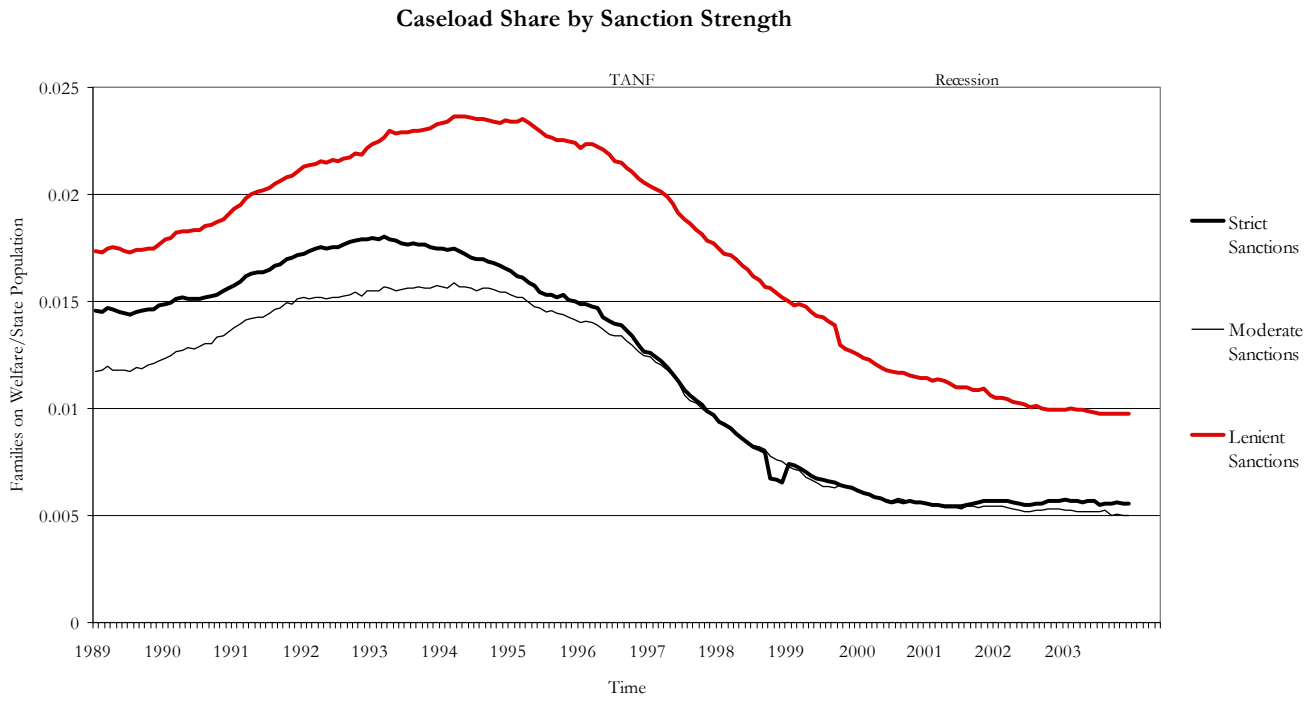


Figure 5 plots the same information but states are organized by sanction strength. Lenient sanction states ("lenient states") have always had much higher caseload shares than both moderate and strict states. Strict states tend to be stricter than lenient states in all areas, and vice versa (see table 1). Therefore, states that chose lenient sanctions after TANF may have always had more lenient, generous welfare programs and, thus higher caseloads. Caseloads fell similarly across all states over the 1990's, but since the recession it appears that the gap between lenient and other sanctions is closing. Without controlling for any other effects, strict states have experienced a slight increase in participation over this time, while lenient states show a small decrease.

Figure 6

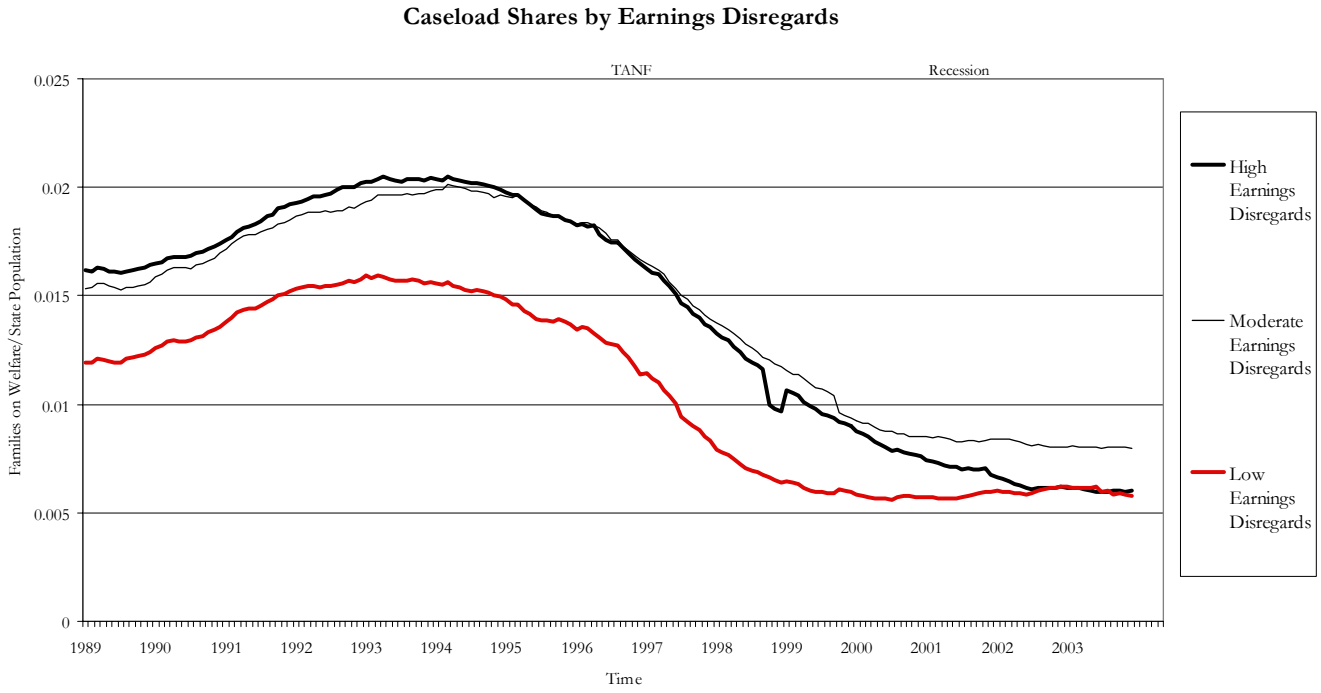


Figure 5 organizes states by earnings disregards. Low earnings disregards states ("low states"), which are more likely to have stricter programs, generally have lower caseloads than other states. Immediately following TANF's passage, caseloads fell particularly quickly in low states but tapered off before the recession and have risen since. Moderate states show a similar but less pronounced trend. High states, however, show steady declines through the recession and at the end of 2003 had the same caseload share as low disregard states.

This initial evidence shows that work incentives clearly matter to caseload changes. It seems that generous earnings disregards fueled the effectiveness of strong work incentive TANF programs after the recession. Recall that in order to have strong work incentives according to Blank and Schmidt (2000) a state needs at least one "strong" parameter and no "weak" ones. About half of strong work incentive states had strict sanctions and about the same number had high earnings disregards. Because the

strict sanction states show slight caseload increases during the recession, but high earnings disregards states continued to shed cases, it may be that high earnings disregards are the driving factor behind the post-recession caseload reductions. None of this preliminary analysis, however, accounts for factors that would affect caseloads differently across states. For this I turn to more sophisticated statistical techniques.

V.b. Regression Analysis –

Table 3 compares Wallace and Blank's (1999) results (p. 78, table 4 part A) to a replication of their model using my data. The main difference between models 1 and 2 is the date range—I use data from 1989-1998 while Wallace and Blank's are from 1977-1998

Table 3 – Wallace and Blank Replica

Dependent Variable is the change in the natural log of the number of families on welfare divided by population

Variable	Model 1		Model 2	
	Wallace and Blank (1999)		1989-1998	
	12 Lags	24 Lags	12 Lags	24 Lags
Σ Unemployment	0.026	0.04	0.034	0.046
	***	***	(7.37)***	(4.21)***
Σ Waiver	-0.079	-0.138	0.003	-0.073
	***	***	(1.56)	(2.16)**
Σ TANF	-0.277	-0.347	-0.244	-0.399
	***	***	(34.23)***	(25.82)***
SSR	--	--	2.608	2.454
Adj-R^2	--	--	0.2439	0.2016
N	--	--	5763	5763
k	--	--	597	627

For sums, partial F-statistics are in parentheses, for single coefficients, t-statistics are in parentheses

*** indicates significance at the 1% level, ** at the 5% level and * at the 10% level

The results for model 2 are very similar to Wallace and Blank's, and where they differ is easy to account for. First, unemployment coefficients are slightly higher in model 2. Before 1989 caseloads did not follow unemployment as closely as they did between 1989 and 2001 (see Figure 1). Therefore, Wallace and Blank estimate unemployment coefficients from data that is less dominated by as close an

unemployment-caseload relationship as the data in model 2. This implies that their coefficient estimates should be smaller, which they are. Wallace and Blank find that waivers reduce caseloads more than in model 2. Unlike in model 2, they do not turn waivers "off" when TANF is implemented. Therefore, their waiver variables pick up some of the post-TANF caseload decline. This also explains why the TANF estimates with 24 lags in model 2 are larger than Wallace and Blank's. Overall, Table 3 shows that my data approximate Wallace and Blank's results quite well.

Table 4 gradually estimates equation 5, by adding wage and benefit information, undifferenced policy dummies and additional dates to model 2. Model 3 adds the ratio of 20^{th} percentile wage to maximum welfare benefit for a family of three. This variable is statistically significant with both lag lengths, but fails to improve the overall fit of the model. A further complication is that when 24 lags of unemployment and policy are included, the wage-benefit ratio switches signs but remains significant. This coefficient should be negative because when the wage-benefit ratio rises, wages become more attractive relative to benefits and caseloads should fall. The perverse sign may be a function of the first differences. Excluding states one at a time from model 3 shows that the lowest benefit states cause the sign switch. It is possible that behavioral responses differ at extreme values of this ratio. The cross-sectional variation which reflects these extremes is removed by differencing, though, and may alter the coefficient on the wage-benefit ratio. To follow theory, the wage-benefit ratio is included in all subsequent models, and other coefficient estimates are unchanged when it is included.

Table 4 – Initial Estimates

Dependent Variable is the change in the natural log of the number of families on welfare divided by population					
Variable	Model 3		Model 4	Model 5	Model 6
	1989-1998		1989-1998	1989-2001	1989-2003
	12 Lags	24 Lags	24 Lags	24 lags	24 Lags
Σ Unemployment	0.036	0.048	0.042	0.041	0.071
	(7.52)***	(8.49)***	(3.48)***	(2.92)***	(7.75)***
Σ Waiver	0.004	-0.069	--	--	--
	(1.52)	(4.04)***			
Σ TANF	-0.241	-0.393	--	--	--
	(32.23)***	(47.97)***			
Σ Wage to Benefit Ratio	-0.015	0.0002	-0.010	0.002	0.003
(6 periods)	(16.49)***	(15.94)***	(15.92)***	(10.53)***	(71.52)***
Waiver	--	--	-0.008	-0.008	-0.007
			(-8.63)***	(-7.51)***	(-6.00)***
TANF	--	--	-0.023	-0.017	-0.011
			(-27.1)***	(-23.54)***	(-16.96)***
SSR	2.555	2.406	2.427	4.236	6.909
Adj-R^2	0.1512	0.1958	0.1953	0.1325	0.0897
N	5508	5508	5508	6881	8870
k	603	633	593	593	593

For sums, partial F-statistics are in parentheses, for single coefficients, t-statistics are in parentheses
*** indicates significance at the 1% level, ** at the 5% level and * at the 10% level

Models 4-6 replace Wallace and Blank's policy lags with undifferenced waiver and TANF dummies. Again, these coefficients represent the change in trend in caseloads when waivers or TANF were in place. Model 4 uses Wallace and Blank's time period, model 5 includes all dates before the recession and model 6 includes all dates. The total estimated effect of TANF over the 18 months in which it was effect in model 4 is -41.4% (-.023/month * 18 months)—nearly identical to the estimates using implementation lags. Consistent with the literature, the per-month effect of waivers is about a third of TANF's. The similarity of these estimates means that undifferenced policy dummies are a reasonable approximation of implementation lags.

Adding dates reduces the per-month TANF effects, as expected. Because this dummy structure imposes the same caseload change to each month a policy was in place, adding months during which caseload declines slowed means that the coefficient per month should be lower. TANF's total estimated effect on caseloads in model 5 and model 6 are -76.5% (-0.017/month * 46 months) and -91.3% (-0.011/month * 84 months), clearly too high. The models explain less and less of the variation in caseload change as more dates are added, signaling that while TANF was important in caseload changes over the whole time period, its effects have changed since implementation. That these estimates of policy effects are inaccurate in their total magnitude does not detract from basic conclusions: TANF had an important caseload reducing effect relative to AFDC even considering the 2001 recession.

Table 5 examines policy effects in more detail by accounting for the program parameters in figures 3-5. Rather than a single TANF dummy, model 8 includes dummies for the strong, mixed and weak work incentive TANF programs shown in figure 3. Unlike figure 3, these estimates control for the caseload effects of unemployment (24 lags), wages, benefits, waivers and other unobserved factors that affect caseloads. The strong, mixed and weak coefficients are comparable to the TANF coefficients in model 6.

Model 8 shows that strong and mixed work incentive TANF programs have reduced caseloads more than have weak work incentive TANF programs. While all types of programs affected caseloads, strong and mixed states had significantly greater effects than weak incentive states.

Table 5 – Policy Categories

Dependent Variable is the change in the natural log of the number of families on welfare divided by population
From June 1989-December 2003

Variable		Model 8	Model 9
Σ Unemployment		0.071	0.073
		(7.82)***	(8.28)***
Σ Wage to Benefit Ratio		0.003	0.005
		(11.94)***	(12.20)***
Waiver		-0.007	--
		(-5.99)***	
TANF Programs			
Strong Work Incentives		-0.012	--
		(-12.31)***	
Mixed Work Incentives		-0.011	--
		(-14.82)***	
Weak Work Incentives		-0.008	--
		(-7.39)***	
TANF Policies			
Earnings Disregards			
	High	--	-0.004
			(-1.97)**
	Medium	--	-0.002
			(-1.12)
	Low	--	-0.003
			(-1.8)*
Sanctions			
	Strict	--	-0.006
			(-2.93)***
	Moderate	--	-0.003
			(-1.82)*
	Lenient	--	-0.005
			(-2.93)***
Time Limits			
	Strict	--	-0.001
			(-0.35)
	Moderate	--	-0.003
			(-1.47)
	Lenient	--	-0.003
			(-1.58)
SSR		6.901	6.914
Adj-R^2		0.0905	0.0883
N		8870	8870
k		595	600

For sums, partial F-statistics are in parentheses, for single coefficients, t-statistics are in parentheses
*** indicates significance at the 1% level, ** at the 5% level and * at the 10% level

Model 9 includes the specific policy dummies discussed in section II. Like the undifferenced waiver and TANF dummies these coefficients measure the change per month that each kind of policy was in place. Achieving statistical significance with this kind of specification is difficult for several reasons. First, the cross sectional component of these data, 51 states, is relatively small. Policies do not vary enough across states to allow precise estimation of their separate effects (see table 1). Moreover, policies may combine in powerful ways to make the effect of one program parameter a partial function of that state's other program choices. High earnings disregards, by encouraging employment but increasing time on the welfare rolls, may work at cross purposes with strict time limits, which create explicit incentives to leave welfare.[21] Another problem is that policies are simply not measured well, so this policy specification leaves out a great deal of nuance (like the difference between receipt time limits or work requirement time limits, or whether or not recipients disregard a percentage of earnings or a certain dollar amount). It also contains virtually no change in state policy over time; a useful natural experiment.

That said, model 9 does yield some reliable conclusions. Earnings disregards and sanctions reduce caseloads, but in both cases the distinction between levels of severity is unclear. Their magnitudes are between a third and a half of the TANF dummy in model 6. Time limits appear not to have any effect on caseloads even though all time limits were binding by 2003. Table 5 provides more detail on the role of policy, but it may not be entirely accurate because it forces these policies to take the same per-month effect in both boom and recession.

[21] Note that this problem is largely avoided in the work incentive specifications because they consider all of a state's policy decisions together rather than trying to separate specific effects.

Table 6 presents two sets of estimates comparable to models 6 and 8,that allow policy effects to differ in the boom and bust. Model 10 includes a dummy for TANF, a dummy for the 90s boom, and an interaction between these two. The TANF*boom dummy gives TANF's extra effect on the caseload trend in the boom. The TANF dummy gives the change in trend attributed to TANF in the recession. Model 11 interacts the work incentive variables with dummies for the national period of falling and rising unemployment rates. It is important to note that estimating the effect of TANF in the bust requires model 10's dummy structure, because with a single TANF dummy contains no cross-sectional variation after all states had implemented the program. Therefore, a TANF*bust dummy would be indistinguishable from a recession dummy. When more than one kind of policy is considered, though, cross-sectional variation allows the estimation of relative effects so the more straightforward boom-bust interaction is acceptable. These models improve on most earlier research by including data from the recession and allowing the effect of policy to differ between the boom and the recession. They also provide more accurate per-month estimates for each than in table 5.

Model 10 shows that TANF's per month effect in the recession (-0.008) was almost half of its effect during the boom (-0.008 – 0.005 = -0.013). This is to be expected, of course, both because any work-focused program will be less effective in a recession and because caseload declines have an obvious lower bound so they cannot fall forever. Importantly, though, TANF still reduces caseloads in the recession.

Model 11 uses this specification with work incentive programs. Strong and mixed work incentives have significantly negative effects on caseloads in both boom and recession. These effects are nearly identical to those estimated for TANF in model 10.

Weak work incentives are also associated with caseload reduction in the boom, but the effect is smaller than in other states. In the recession, this effect is not statistically different from zero. This provides stronger evidence that the incentives embedded in policy matter. While no state had trouble reducing caseloads over the 90s, Table 6 shows that during recession caseloads are affected by different levels of policy.

Table 6 – Falling vs. Rising Unemployment Models

Dependent Variable is the change in the natural log of the number of families on welfare divided by population
From June 1989-December 2003

	Model 10	Model 11
Σ Unemployment	0.026	0.041
	(1.42)*	(2.98)***
Σ Wage to Benefit Ratio	0.003	0.009
(6 periods)	(2.92)***	(12.09)***
Waiver	-0.005	-0.008
	(-3.93)***	(-6.81)***
TANF		
TANF*Boom	-0.005	--
	(-4.05)***	
TANF in Bust	-0.008	--
	(-8.65)***	
TANF Programs		
Strong Work Incentives		
Boom	--	-0.018
		(-14.24)***
Bust	--	-0.005
		(-3.55)***
Mixed Work Incentives		
Boom	--	-0.017
		(-17.44)***
Bust	--	-0.005
		(-5.07)***
Weak Work Incentives		
Boom	--	-0.014
		(-9.9)***
Bust	--	-0.001
		(-0.8)
SSR	6.719	6.766
Adj-R^2	0.115	0.108
N	8870	8870
k	595	598

For sums, partial F-statistics are in parentheses, for single coefficients, t-statistics are in parentheses
*** indicates significance at the 1% level, ** at the 5% level and * at the 10% level

The final specification, in Table 7, applies the boom-bust dummy structure to the program parameters estimated in model 9. There are two main results in model 12. First, while medium and high earnings disregards are not associated with caseload reductions during the boom, high earnings disregards do have a significant negative effect on caseloads in the recession. In fact, they are the only policy that has a statistically significant effect during the recession. This is an important conclusion because it signals that in high earnings disregards states, a long-run caseload reducing scenario may have been reached. That is, generous financial incentives may have increased labor force participation enough that 6-7 years after reform, low-income women may not need to use welfare cyclically.

The second result is that sanctions appear not to have any effect on caseload change in the recession, while they have the greatest and most precisely estimated effects in the boom. This is a departure from the literature, which finds sanctions to reliably reduce caseloads. Taken together, these two results imply that negative incentives to work are less effective at reducing caseloads in the recession than in the boom and positive incentives to work are more effective.

The effects of time limits follow the same pattern as sanctions, but these results are less reliable. Most time limits became binding near beginning of the recession, so the boom-bust specification misrepresents the mechanical effects of time limits. If the hit date came before the recession, the boom coefficient may be lower, and if the hit date came after the recession the bust coefficient may be lower. Mechanical reductions due to time limits are unrelated to the recession, however, so the estimates of time limit effects are not entirely reliable. In general, model 12 shows that stricter time limits had greater

negative effects on caseloads in the boom, but that no time limit had an effect in the recession.

Table 7 –Policy Choices in Boom vs. Bust

Dependent Variable is the change in the natural log of the number of families on welfare divided by population

Variable		Model 12 1989-2003		
Σ Unemployment		0.046		
		(3.49)***		
Σ Wage to Benefit Ratio		0.011		
(6 periods)		(11.67)***		
Policies (Waiver and TANF)				
Earnings Disregards		HIGH	MEDIUM	LOW
	Boom	-0.003	-0.002	-0.006
		(-1.39)	(-1.08)	(-2.86)**
	Bust	-0.003	-0.001	--
		(-1.95)*	(-0.34)	
Sanctions		STRICT	MODERATE	LENIENT
	Boom	-0.01	-0.005	-0.006
		(-4.65)***	(-2.69)***	(-3.27)***
	Bust	0.004	0.004	0.000
		(0.72)	(0.74)	(0.07)
Time Limits		STRICT	MODERATE	LENIENT
	Boom	-0.005	-0.006	-0.002
		(-2.42)**	(-2.82)***	(-0.91)
	Bust	-0.006	-0.005	-0.005
		(-1.15)	(-1.02)	(-0.87)
SSR		6.751		
Adj-R^2		0.1089		
N		8870		
k		608		

For sums, partial F-statistics are in parentheses, for single coefficients, t-statistics are in parentheses
*** indicates significance at the 1% level, ** at the 5% level and * at the 10% level

The preceding models estimate monthly changes in the natural log of a caseload to population ratio. This is several steps away from what policy-makers care about: caseload levels. I now briefly examine how well these models predict caseload levels.

V.c. Simulations -

Figure 7 plots predicted national caseload levels against actual levels for model 6, the simplest specification that uses all dates and undifferenced dummies. This model

does not account for different policy choices across states nor does it allow policies to have different per month effects in the boom and recession. This model under-predicts caseloads before TANF and begins over-predicting shortly after 1996. The tapering off of caseloads near the recession appears slightly in these predictions, but in general predicted caseloads are too linear around this time.

Figure 7

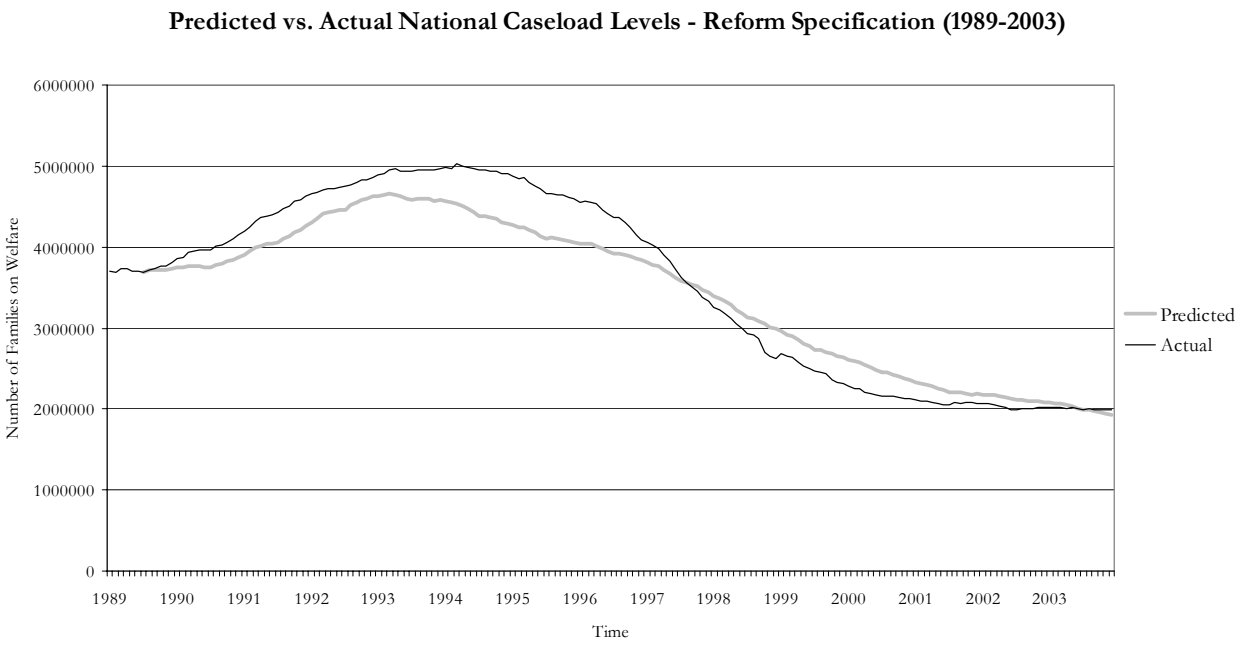
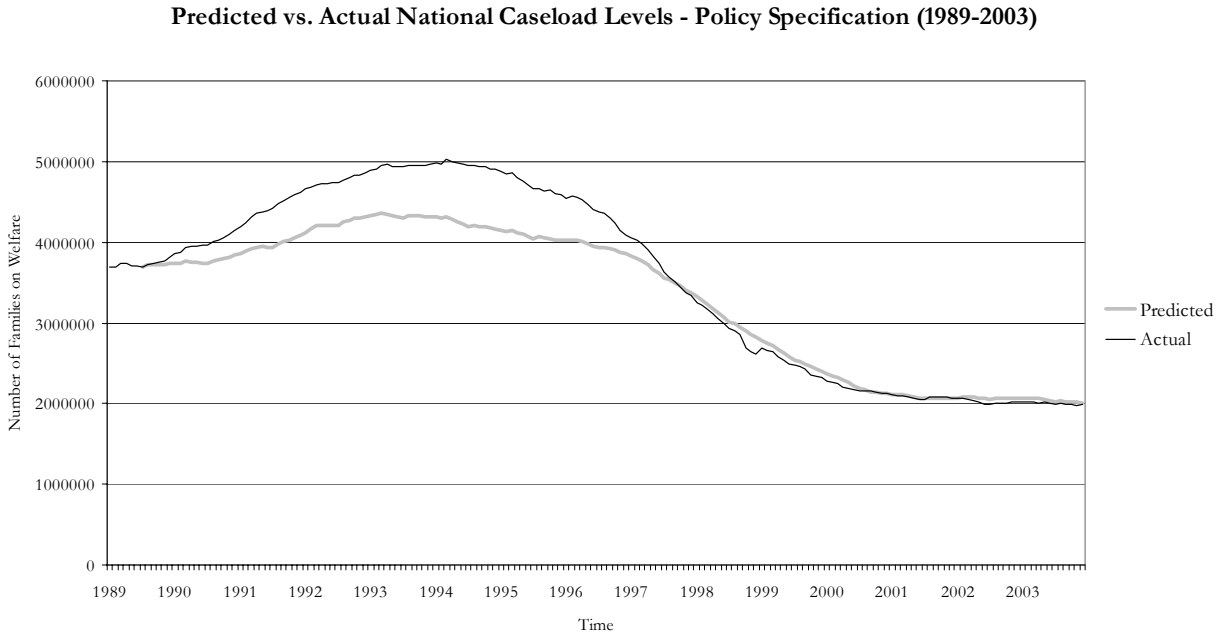


Figure 8 shows predicted values for model 12, the most detailed with regard to policy and the recession. This model characterizes states' policies by parameters and by the severity of those parameters. Model 12 also allows policy effects to differ between boom and recession. While pre-TANF under-prediction is worse, this model does a much better job of predicting caseload levels since TANF's passage and especially since the recession. The two series are virtually identical from 1997 on not only in shape, but in their level. This lends credibility to the estimated policy effects in model 12.

Figure 8



VI. Conclusions –

This paper finds that since the recession, welfare reform reduced the rate of program participation, but that specific policy choices matter to caseload changes during this time. Policy combinations that more strongly encourage work move women off of welfare better than those that provide weaker incentives. Specifically, sanctions appear not to have any impact on caseloads in the recession, while they were quite effective at reducing caseloads in the 90s. Generous earnings disregards, on the other hand, do have a significant downward effect on caseloads in the recession, while they were not statistically important in the boom. Given that high earnings disregards increase eligibility and, therefore, participation in the short-run, this finding implies that the caseload reducing effect of increased labor market attachment for AFDC recipients who did not work is quite strong.

The earnings disregard result is puzzling, though, because high earnings disregards had their negative effect in the time since the recession—not when one would expect the labor market to be drawing women off welfare. There are two possible explanations for this. It could be that the increased labor market contact over the 5-7 years since reform allowed women in generous disregards states to gain labor market skills, enough wage advancement, or sufficient job prestige to weather a recession and avoid or even leave welfare. Alternatively, it could be that generous disregard states tend not to have chosen the strict policies that led to the most rapid caseload declines in the boom. As such, their caseloads had not yet reached some hypothetical lower limit in 2001, while stricter policy states' caseloads had. If this is the case, then by isolating dates after March 2001, my specifications find that certain policies have reduced caseloads in recession while, in fact, these policies simply reduced them more slowly over the 90's and, therefore, could continue doing so, regardless of the macroeconomy, after stricter states caseloads "bottomed out." While the second explanation may carry some weight it does not erode the policy conclusions because the recession is not a random cut-off date. Unemployment rates did rise after March 2001, so the policy effects are still credible because they were influenced by economic conditions in both periods.

Two main caveats are in order. First, it is not clear that the most recent recession has had its full effect on welfare caseloads. These specifications allow for two years' worth of unemployment lags, and if decisions about welfare come near the end of those two years, then low-income women may still be responding to the jump in unemployment. It will also be interesting to see what has happened to caseloads in the "jobless" recovery that was so prominent in the 2004 presidential campaign. For whom

was it jobless? How were recipients and ex-recipients affected compared to other recessions?

Second, caseloads do not tell the whole story of welfare reform, or any anti-poverty policy. For instance, if states drastically reduced eligibility, caseloads would fall but it is not clear that low-income single mothers would be any better off. At best, caseloads may proxy for well-being and at worst they betray real trends in well-being. While these results are a start to analyzing TANF in recession, we need to know more about how the recession and welfare reform policies have affected a broader range of outcomes for low-income families such as labor force participation, work, earnings, consumption, poverty, and family formation.

Sources

Ashenfelter, Orley. "Estimating the Effect of Training Programs on Earnings," *Review of Economic Statistics*, February, 1978, 60(1), 47-57.

Ashenfelter, Orley. "Determining Participation in Income-Tested Social Programs," *Journal of American Statistical Association*, September 1983, 78, pp. 517-25.

Bartik, Timothy J. and Randall W. Eberts."Examining the Effect of Industry Trends and Structure on Welfare Caseloads," in Sheldon Danziger, editor, *Economic Conditions and Welfare Reform*, Kalamazoo, MI: Upjohn Institute for Employment Research, 1999, 119-157.

Blank, Rebecca. "What Causes Public Assistance to Grow?" *Journal of Human Resources*, Winter, 2000, 36(1), 85-118.

Blank, Rebecca. "Evaluating Welfare Reform in the U.S.," *Journal of Economic Literature*, December, 2002, 40(4), 1105-1166.

Blank, Rebecca, David Card and Philip K. Robins. "Financial Incentives for Increasing Work and Income Among Low-Income Families," in Rebecca Blank and David Card, eds., *Finding Jobs: Work and Welfare Reform*, New York, NY: Russell Sage Foundation, 2000, 373-419.

Blank, Rebecca and Lucie Schmidt. "Work, Wages and Welfare," in Rebecca M. Blank and Ron Haskins, eds., *The New World of Welfare,* Washington, D.C.: Brookings Institute, 2001, 70-102.

Bloom, Dan, Mary Farrell and Barbara Fink. "Welfare Time Limits: State Policies, Implementations and Effects on Families," Manpower Demonstration Research Corporation, 2002.

Council of Economic Advisers. "Technical Report: Explaining the Decline in Welfare Receipt, 1993-1996," Washington, D.C.: Executive Office of the President, April, 1997.

Council of Economic Advisers. "Economic Expansion, Welfare Reform, and the Decline in Welfare Caseloads: An Update. Technical Report," Washington, D.C.: Executive Office of the President, September, 1999.

Danielson, Caroline and Jacob Alex Klerman. "Why Did the Welfare Caseload Decline?" RAND Working Paper, 2004. Available at: http://www.npc.umich.edu/publications/workingpaper04/index.shtml

Figlio, David N. and James P. Ziliak. 1999. "Welfare Reform, The Business Cycle, and the Decline in AFDC Caseloads," in Sheldon Danziger, editor, *Economic*

Conditions and Welfare Reform, Kalamazoo, MI: Upjohn Institute for Employment Research, 1999, 17-48.

Friedman, Milton. *Capitalism and Freedom,* Chicago: University of Chicago Press, 1962.

Gladden, Tricia and Christopher Taber, "Wage Progression Among Less Skilled Workers," in Rebecca Blank and David Card, eds., *Finding Jobs: Work and Welfare Reform*, New York, NY: Russell Sage Foundation, 2000, 160-192.

Grogger, Jeffrey. "Time Limits and Welfare Use," NBER Working Paper No. 7709, 2000.

Grogger, Jeffrey. "Welfare Transitions in the 1990's. The Economy, Welfare Policy and the EITC," NBER Working Paper No. 9472, January, 2003.

Hoxby, Caroline. "School Choice and School Productivity (or Could School Choice be a Tide that lifts All Boats?)," Prepared for the NBER Conference of the Economics of School Choice, February 2001.

Hoynes, Hilary W. "The Employment, Earnings and Income of Less Skilled Workers Over the Business Cycle." in Rebecca Blank and David Card, eds., *Finding Jobs: Work and Welfare Reform*, New York, NY: Russell Sage Foundation, 2000, 23-71.

Leonisio, Michael. "In Kind Transfers and Work Incentives," *Journal of Labor Economics*, October 1988, 6(4), 515-529.

Levy, Frank. "The Labor Supply of Female Heads, or AFDC Work Incentives Don't Work Too Well," *Journal of Human Resources*, Winter, 1979, 14(1), 76-97.

Moffit, Robert. "Work Incentives in Transfer Programs (Revisited): A Study of the AFDC Program," in Ronald Ehrenberg editor, *Research in Labor economics* vol. 8, Greenwich, CT: JAI Press, 1986b, 389-439.

Moffit, Robert. "Incentive Effects of the U.S. Welfare System," *Journal of Economic Literature*, March 1992, 30(1), 1-61.

Moffit, Robert. "The Effects of Pre-PRWORA waivers on Welfare Caseloads and Female Earnings, Income and Labor Force Behavior," in Sheldon Danziger, editor, *Economic Conditions and Welfare Reform*, Kalamazoo, MI: Upjohn Institute for Employment Research, 1999, 91-118.

Pavetti, LaDonna and Dan Bloom. "Sanctions and Time Limits: State Policies, Their Implementation and Outcomes for Families." in Rebecca M. Blank and Ron Haskins, eds., *The New World of Welfare,* Washington, D.C.: Brookings Institute, 2001, 245-269.

Schoeni, Robert and Rebecca Blank. . "What Has Welfare Reform Accomplished?" NBER Working Paper No. 7627, 2000.

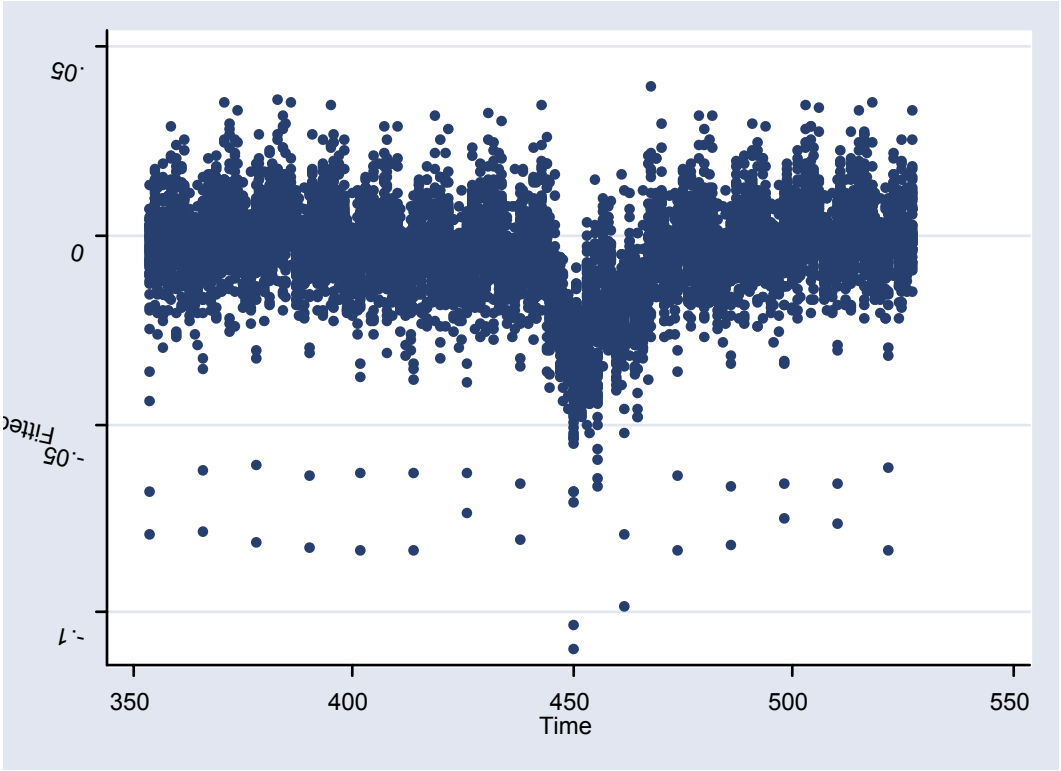
US House of Representatives. Green Book. 2004. Available at http://www.gpoaccess.gov/wmprints/green/.

Wallace, Geoffrey and Rebecca Blank. "What Goes Up Must Come Down? Explaining Recent Changes in Public Assistance Caseloads," in Sheldon Danziger, editor, *Economic Conditions and Welfare Reform*, Kalamazoo, MI: Upjohn Institute for Employment Research, 1999, 49-89.

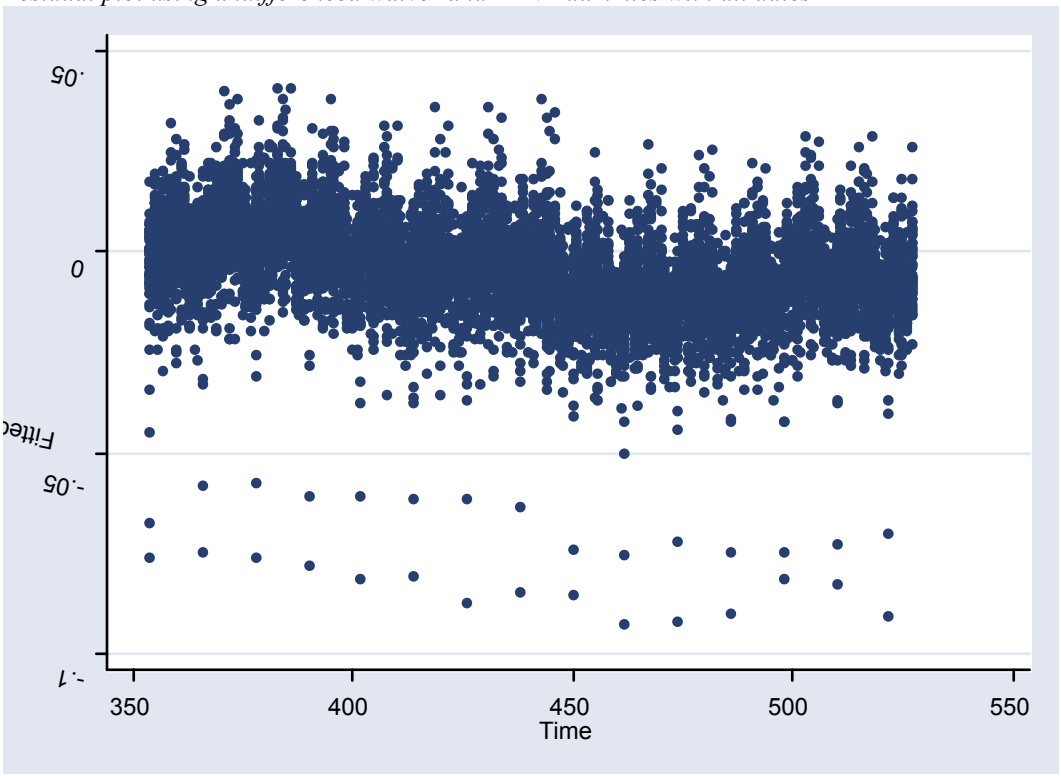
Ziliak, James P., David N. Figlio, Elizabeth E. Davis and Laura S. Connolly. "Accounting for the Decline in AFDC Caseloads." *Journal of Human Resources*. Summer, 20001, 35(3), 570-86.

DATA APPENDIX -

Figure 1 – Residual Plot using Wallace and Blank's 24-lag specification with all dates



Residual plot using undifferenced waiver and TANF dummies with all dates



Note the dip in residuals in Wallace and Blank's residuals when the TANF dummy turns off