Confounded? Welfare Reform and the Earned Income Tax Credit in the 1990s*

Adam Looney†

October, 2024

Abstract

A large literature credits the 1993 Earned Income Tax Credit (EITC) expansion with increasing labor supply and reducing welfare use among single mothers in the 1990s. A concern is that the estimated effects of the EITC are confounded by contemporaneous changes in welfare policy. This paper proposes falsification tests for the unconfoundedness assumption in difference-in-difference (DiD) estimators, which show that standard estimates of effect of the 1993 EITC expansion are spurious, and that the identifying assumptions of the DiD estimator fail to hold. Controlling for the confounding effects of welfare reform identified by these tests, I find little role for the EITC in increasing labor supply in the 1990s.

Keywords: Earned Income Tax Credit (EITC), Labor Supply, Welfare Reform, Differences-in-Differences
JEL Classification: I38, H24, J22

*I thank seminar participants at the University of Chicago, the National Tax Association, IIPF, the Brookings Institution, and the University of Utah, and numerous conversations with colleagues. This research was not funded by any outside organization and the author has no financial conflicts to report.

†David Eccles School of Business, University of Utah, and The Brookings Institution

1 Introduction

Research credits the 1993 expansion of the Earned Income Tax Credit (EITC) for the large and persistent increase in labor supply and the decline in welfare use among single mothers in the 1990s. This evidence from the 1993 expansion represents a substantial share of the evidence regarding the EITC's labor supply effects (Nichols and Rothstein (2015)). And the 1993 expansion is the basis for much of the evidence on the effects of the EITC on other outcomes, such as earnings, income, poverty, self-employment, health, and human capital. Supported, in part, by the view that the labor supply changes in the 1990s reflected the pull of "carrots" rather than the "sticks" of welfare policies, "virtually all gains in spending on the social safety net for children since 1990 have gone to families with earnings" (Hoynes and Schanzenbach (2018)). The evidence gleaned from the 1990 experience continues to inform social policy today.

A concern in the analysis of the 1993 EITC expansion is that it is confounded by contemporaneous changes in cash welfare associated with "welfare reform" (see, e.g. Mead (2014), Kleven (2024)). While this threat to identification is known, it is not clear whether or to what extent it biases the estimated effect of the EITC. Estimates of the effect of the EITC in the literature assume unconfoundeness conditional on observable characteristics of sample individuals and economic conditions, and this assumption is supported by the standard assessment of difference-in-difference (DiD) estimators, which show that pre-reform trends in treatment and control groups are "parallel." However, the available econometric tools provide little guidance how to assess whether or not the DiD estimator is counfounded. As a result, the debate about whether the estimated effect of the 1993 EITC expansion is confounded remains unresolved.

To answer this question, I implement falsification tests that examine whether the canonical difference-in-difference (DD) estimators of the effect of the EITC in the published literature are biased. The first test assumes confoundedness is bias induced by an omitted variable, and uses placebo tests of the DiD estimator to identify any bias, its magnitude, and its source.

Placebo tests ask whether a empirical analysis produces treatment effects even in the absence of actual treatment and, if so, can reveal why. In the EITC literature, the DiD estimate is formed from the comparison of a treatment group (e.g. single mothers with two or more children, for whom the EITC was increased the most) and one or more control groups (e.g. mothers of one child or childless women), before and after the EITC was expanded. The placebo tests compare “treatment” and “control” groups composed of equally-EITC-treated individuals, and re-estimate the DiD estimators. By construction, the resulting treatment effects should be zero (ignoring sampling error). Instead, these placebo tests produce large, spurious “effects” every time two groups with different pre-reform rates of welfare use are compared, in direct proportion to the difference in their pre-reform welfare use, suggesting exposure to welfare reform is an omitted variable. Moreover, the spurious placebo treatment effects and actual treatment effects are closely related, suggesting the bias is large.

In the second analysis, I assume confoundedness arises from imbalances in the composition of the treatment and control groups. Starting with the equation for the DiD estimator, I use decomposition methods to express the estimator as the sum of the causal effect of the treatment on the treatment group plus a bias term equal to the conditional time trends of the control group weighed by the difference in composition between treatment and control group. Intuitively, this bias term arises when information available exclusively from changes in one group implies a predictable divergence in outcomes between treatment and control group because of compositional differences—in other words, non-parallel trends; the identifying assumption in the standard DiD estimator is that this bias term is zero.

To examine this, I compare individuals with different pre-reform rates of welfare use within each EITC treatment and control group. First, this analysis shows that individuals with higher pre-reform rates of welfare use experience greater increases in employment in all groups. Second, individuals with higher pre-reform rates of welfare use are over-represented in the treatment group. As a result, even in the absence of treatment, the group outcomes would not be parallel, suggesting that differences in the composition of the two groups bias the DiD estimator. Using the traditional graphical comparison of trends between treated and untreated groups used to motivate the DiD analysis, I show that the changes over time can be entirely explained by differences in the composition of treatment and control groups rather than by the treatment effect of the EITC.

This analysis suggests that exposure to welfare reform—proxied, for example, by time effects proportionate to predicted welfare use—is an important confounding factor. I replicate several canonical DiD estimators with controls for these confounding effects, and find no economically or statistically significant role of the 1993 EITC expansion in increasing labor supply in the 1990s.

As an alternative strategy to identify confounding factors ex-ante, I use off-the-shelf machine learning (ML) models for covariate selection. These models help form hypotheses about potential confounding factors and, in the present case, readily identify the same time-varying covariates described above, like measures of exposure to welfare reform. In addition, the predicted outcomes of these models can provide an intuitive graphical assessment of whether the parallel trends assumption is plausible. In the EITC context, predictions trained on one group's data (e.g. the control group) and out-of-sample predicted for another group (e.g. the treatment group) suggest that the parallel trends assumption should not be expected to hold, leaving little room for the effect of the 1993 EITC expansion on labor supply.

Taken together, these results suggest that the changes in labor supply were primarily caused by the sweeping changes in welfare policy that culminated in "welfare reform" in 1996 rather than the EITC. Ethnographic histories of the welfare reform era (e.g. DeParle (2004) Weaver (2000)) point to the role of "caseload reduction" or "changes in the 'culture' of welfare receipt, including the behaviors and attitudes of recipients and caseworkers" (Levine and Whitmore (1997)) as the most salient and important cause of the reduction in welfare use and rise in employment among single mothers in the 1990s. Because exposure to these policies differed across groups, these policies caused disproportionate changes in the levels of employment and welfare use among groups with high rates of welfare use–like mothers with more children. In typical DiD estimators, these changes are erroneously attributed to the EITC rather than welfare reform.

This paper proceeds as follows: The background section provides historical context on changes in welfare use and employment in the 1990s, and the econometric evidence regarding the effects of the 1993 EITC expansion. Section three describes the methodology regarding the falsification tests and how differences in the composition of treatment and control groups may bias the estimator. The next section (4) presents the results of these tests, estimates models that control for confounding covariates, and illustrates how machine learning methods can help identify potential sources of bias ex-ante. Section five concludes with a discussion of the implications of this analysis for understanding the welfare reform era and contemporary debates regarding social policy.

2 Background

Prior to the mid-1990s the employment rate of single women exceeded that of single mothers by an average of about 20 percentage points (Figure 1). That gap was entirely eliminated over the latter half of the 1990s. At the same time, the number of women receiving welfare fell by more than two thirds.

A large literature finds that the 1993 EITC expansion was an important cause of this surge in employment and decline in welfare (see e.g. Blank (2002), Hotz and Scholz (2001), and Nichols and Rothstein (2015) for comprehensive reviews of accumulated evidence).[1] More recently, Kleven (2024) provides a reappraisal of the EITC literature in which he argues that the evidence from the 1993 expansion should be discarded because it is confounded by the effects of welfare reform. In his critique, he raises several "puzzles" like the "fanning-

[1]The literature on the behavioral response to the EITC originates with Eissa and Liebman (1996)'s analysis of the 1986 expansion, and includes a large literature examining earlier reforms (e.g. the 1975 expansion (Bastian (2020)). These studies do not focus on the 1990s welfare reform era, and my analysis does not assess these policy changes.

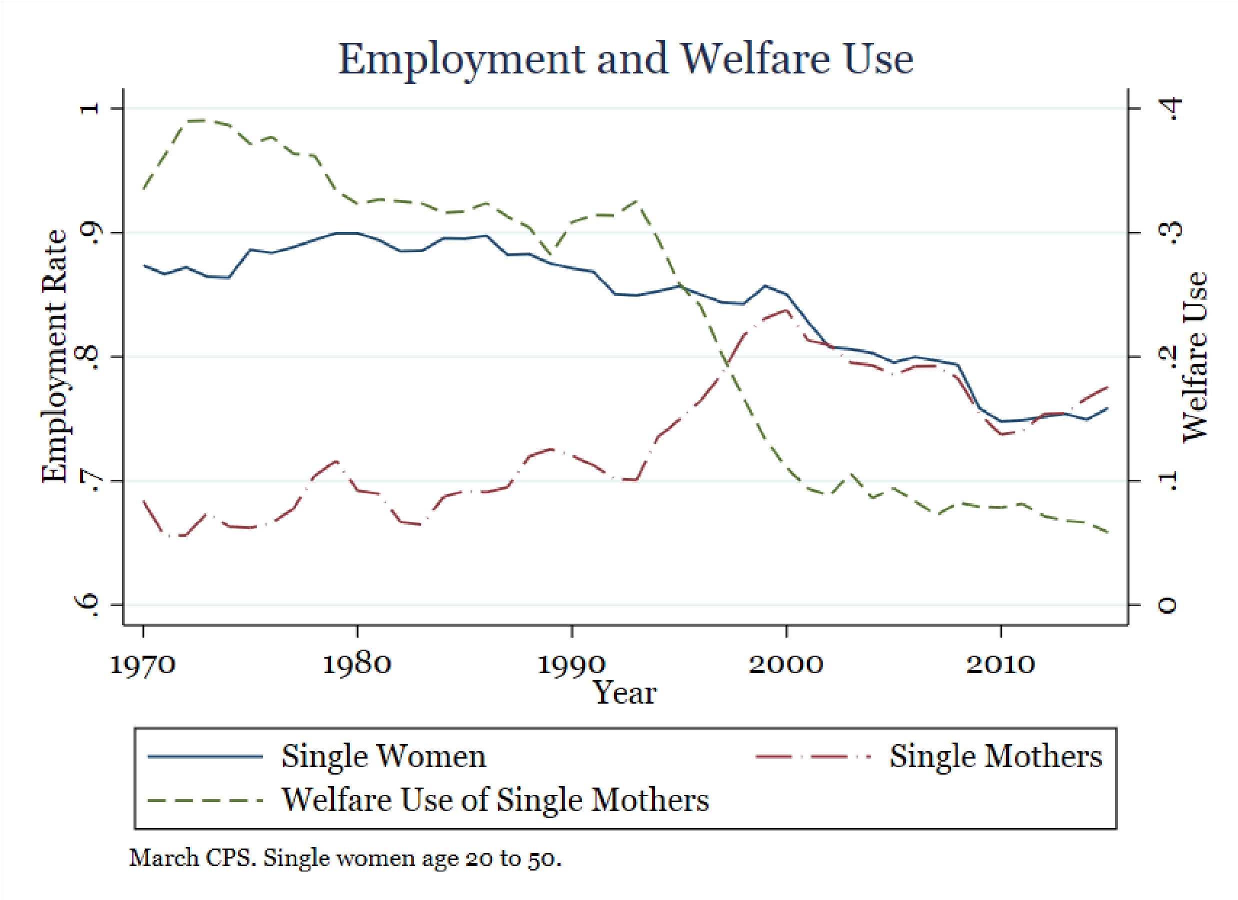

Figure 1: Source: March CPS (Flood et al. (2024))

out" of employment effects by family size and differences in employment across groups with different rates of pre-reform welfare use, which he argues arises because of welfare reform policies. However, it is not clear whether or how much these factors bias the estimated effect of the EITC, or how to address the bias.

The 1993 expansion of the EITC was the largest EITC expansion and substantially increased benefits for low-income working families, especially those with children. The maximum EITC for families with two or more children nearly doubled from \$2,610 in 1993 to \$56,49 in 1996 (in 2021 dollars); for families with one child, the maximum EITC rose from \$2,477 to \$3,418; and a new benefit was created for childless individuals of about \$520. In addition, the phase-in rate—the percentage of earned income that determines the credit amount—was increased (e.g. from 19.5% to 40% for families with two or more children), and the income level at which the credit began to phase out was raised, allowing more families to qualify for the maximum credit.

The primary evidence of the effect of the EITC is derived from DiD estimators, which compare mothers who were eligible for expanded EITC benefits (e.g. mothers of 2+ children) to less treated individuals (mothers of one child, who received a smaller EITC increase) and to untreated (or less treated) individuals (such as single women eligible for only a small EITC). In this strategy, the basic estimating equation has the following form:

$$y_{it} = \beta_0 + \beta_1 \text{after}_t + \beta_2 \text{treat}_i + \beta_3 (\text{treat}_i \times \text{after}_i) + X\beta + \epsilon_{it} \tag{1}$$

Where y_{it} is the outcome of interest (e.g., employment, earnings, or welfare use); treat_i is an indicator (or a vector of indicators) for whether the individual is in a treatment group—in this case, an indicator variable for the number of EITC-eligible children (0, 1, or 2+); after_t is an indicator for time (pre- or post-treatment) or, more typically, a vector of individual year effects; and X_i is a vector of individual characteristics. In addition to individual characteristics, some specifications include controls for certain time-varying policies, such as welfare reform, state welfare waivers, state economic conditions, or state generosity of welfare programs, either directly (e.g., as a welfare reform dummy variable indicating that welfare reform has been implemented in a state in a given year) or by including state effects or state-by-year effects.

In some specifications, treat_i is an indicator for being in the treatment group. Alternatively, it is a simulated dollar value of the EITC (or the net-of-EITC tax burden) which corresponds to the expected amount available to the average single mother in each treatment or control group. In either case, this model captures the effect of the EITC with the coefficient β_3, which measures the difference between the treatment and control group after the implementation of the reform.

As illustrative examples, I replicate the estimating equations of four important papers in the literature that use variations of this empirical strategy to identify the effect of the 1993 EITC. I focus on these papers because they are representative of the published evidence on the EITC and use common and well-known data from the March CPS, but use different parameterizations of the EITC, examine the EITC's effects on different subpopulations, and differ in whether they incorporate federal or federal and state EITC expansions.

These papers are: (1) Strain and Schanzenbach (2021), which in which the EITC expansion is represented as a dummy variable that is equal to zero before 1994, equal to one for mothers with one and two children after 1995; 0.92 for mothers of one child and 0.5 for mothers of two children in 1994 (reflecting the fact that the EITC expansion was partially phased in that year); and equal to one for mothers of one child and 0.78 for mothers of two children in 1995. The policy variation is at the national level. (2) Meyer and Rosenbaum (2001), in which the EITC's effect is incorporated in the variable "taxes if work" or the average level of taxes if working by year and number of children (Table 3). In my replication of these two examples, policy variation in the EITC is at the year-by-number-of children-level (0, 1, or 2+ children). (3) Hoynes and Patel (2018), in which the EITC is represented as a simulated value in each state (including any state-level EITC) by number of children (0 to 3). For each of these three studies, the authors report estimates for subpopulations like mothers with different levels of educational attainment (e.g. less than high school or high school only). (4) Michelmore and Pilkauskas (2021), which examines the effect of the EITC on the labor force participation of mothers with different-aged youngest children. In this analysis, the EITC is modeled as a simulated EITC value in each state and for each number of children (1-3; they exclude mothers without children), and the EITC treatment variable is interacted with group dummies for age of youngest child.

To focus on the 1993 EITC expansion, I estimate representative DiD regressions based on each of the specifications described in the papers above over the period from 1991-2001 in the March CPS. This time period excludes earlier EITC expansions but surrounds the 1993

EITC expansion and welfare reform.[2] I follow, as closely as possible, the sample selection choices that best represent the literature: non-married women aged 20 to 50 (excluding childless women enrolled in school). I exclude single mothers whose youngest child is older than 18. I include covariates typical in the literature: a quartic in the mother's age, dummy variables for education level (less than high school, high school, some college, college graduate), race (white, Black, other), age of youngest child (0-2, 3-5, 6-12, 13+), and marital status. Rather than including controls for state-level policies and local economic characteristics, I include state-by-year fixed effects.[3] In the analysis presented below, the dependent variable is employment last year (defined as having had positive earnings). However, the overall analysis is similar when the dependent variable is welfare use last year, current employment, or annual earnings last year.

Figure 2 summarizes the results of estimating these specifications, and compares the resulting coefficient estimates to those in the published literature.[4] In each specification, the estimated effect of the EITC is economically large and statistically significant. Despite somewhat different time periods, control variables, and treatment variables, these closely resembles published estimates.[5] In other words, these specifications strongly imply that the EITC had large effects on employment. In each of these DiD specifications, the identifying assumption is that the outcomes of treatment and control individuals would have been the same in the absence of the EITC after controlling for time effects and other characteristics of sample individuals and the applicable economic or policy environment (as captured by state by year effects).

A key threat to this identifying assumption is the effect of contemporaneous changes to

[2]Note that the period of analysis in Meyer and Rosenbaum (2001) is from 1984-1996, and specifically excludes post-1996 welfare policy changes. Hoynes and Patel (2018) study 1991 to 1998. In their analysis of the 1993 expansion, Strain and Schanzenbach (2021) examine 1989 to 1998 (when examining the 1993 expansion). Michelmore and Pilkauskas (2021) study the period from 1990 to 2016.

[3]Summary statistics for this well-known dataset are included in Appendix A.

[4]The treatment effects happen to fall on the same scale despite some being dummy variables and others being simulated dollar amounts (in thousands of dollars).

[5]The exception is that including more post-1996 data, the coefficient estimates on "tax if work" is substantially larger than estimated by Meyer and Rosenbaum (2001).

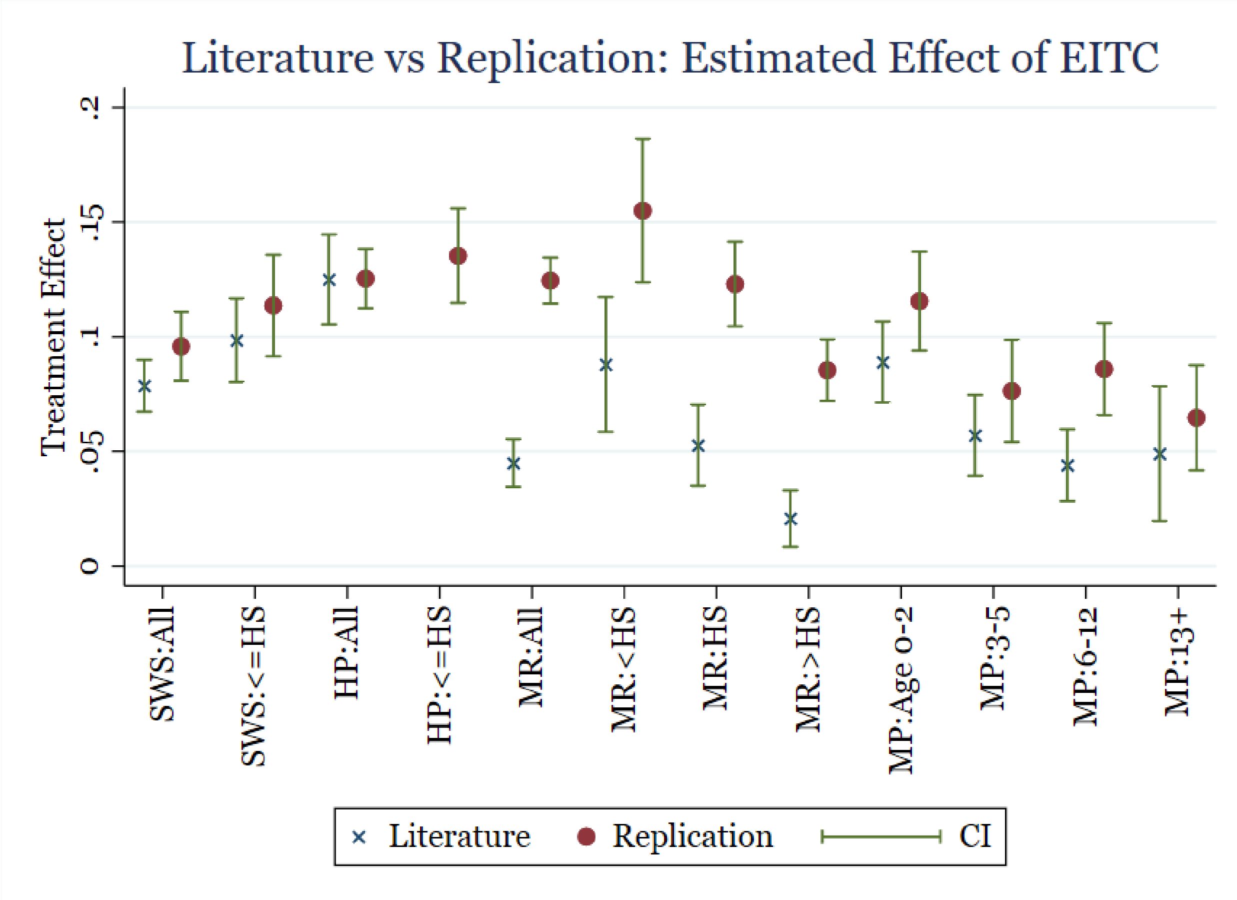

Figure 2: Note: The figure replicates the central specifications of the effect of the EITC of Strain and Schanzenbach (2021) (SWS); Hoynes and Patel (2018) (HP); Meyer and Rosenbaum (2001) (MR); and Michelmore and Pilkauskas (2021) (MP) within a consistent March CPS sample from 1991-2001 within the sub-samples indicated on the x-axis, and compares the replicated estimates to published estimates. (HP does not provide an estimate of the effect of the EITC on a high-school-only population, which is why it is omitted.)

welfare policy ("welfare reform"). Aid to Families with Dependent Children (AFDC) was "ended as we know it" in the words of then-President Clinton, with the passage of the Personal Responsibility And Work Opportunity Reconciliation Act of 1996. Even before 1996, many states had sought and received waivers to change AFDC parameters to incentivize employment or sanction non-compliance with program rules. And some states effected changes in welfare receipt simply through administrative changes, like burdening would-be recipients with onerous paperwork (DeParle (2004)).

Ethnographies of the welfare reform era suggest a major role for changes in bureaucratic processes and attitudes rather than employment subsidies like the EITC. The Rockefeller Institute's implementation study, for example, found that many local sites had reoriented their efforts toward diverting would-be recipients away from welfare and toward work; "local officials and workers view state officials as wanting to see, above all else, lower caseloads" (as quoted in Weaver (2000)). In areas with the most dramatic reductions in welfare use, like Wisconsin, where the number of program recipients declined sharply from 226,197 in 1994 to 96,054 in 1997, ethnographers attributed the changes to work requirements enforced by new bureaucratic procedures (DeParle (2004)).

An alternative hypothesis of the cause of increases in employment in the 1990s based on these historical narratives is that exposure to these bureaucratic changes and new social norms was not equal across the population of single women, but instead was proportional to a person's exposure to the welfare system. Groups of women with higher rates of participation in the early 1990s were "treated" more by welfare reform than those who had little interaction with the welfare system. The econometric concern is that controlling for the average effect of welfare reform, e.g. with year or state-by-year effects or other state-level covariates, does not address the heterogeneity in exposure across individuals.

This is the substance of Kleven (2024)'s concern that the labor supply effects of the EITC in the 1990s are confounded. Pre-reform indicators of welfare use are stronger predictors of the increase in employment over the 1990s than EITC-qualifying number of children (Looney and Manoli (2016); Kleven (2024)).

Indeed, a back-of-the envelope calculation suggests the proportional exposure hypothesis has compelling predictive power. For instance, between 1993 and 2000, the share of single mothers who received welfare and had no employment declined from 20.5 percent to 4.7 percent, a 77 percent reduction. For mothers of one child, the decline was from 12.4 percent to 2.8 percent (a 77 percent reduction); for mothers of two or more children, decline was from 27.0 percent to 6.3 percent (again 77 percent).[6] However, since mothers with differing

[6](The percentage decline is was also approximately 77 percent many other groups of mothers, such as mothers of young children or older children; mothers of three or more children; and mothers with different

numbers of children had different pre-reform levels of welfare use, the 77 percent decline in welfare-reflated non-employment implies a 20.7 percentage point increase in employment among mothers of two or more children but a 9.6 percentage point increase in employment among mothers of one child, a difference-in-difference of 11.1 percentage points. In fact, the raw difference-in-difference in employment between these groups was 11.8 percentage points. The DiD estimate of the EITC's effect on employment is identical to what one might expect if the entire change in employment was caused by removing 77 percent of mothers from welfare at random.[7]

While the potential confounding effect of welfare reform is longstanding, and Kleven's evidence of "unexplained" changes in employment by number of children or by pre-reform rates of welfare use is suggestive of bias, it has failed to persuade researchers either of the existence of a confounding factor, its magnitude, or how to address it, nor that conventional DiD estimators are flawed. For instance, in their rejoinder to Kleven, Strain and Schanzenbach (2021) re-estimate canonical DiD estimators of the effect of the EITC in the 1990s (and around other reforms), make methodological choices that are standard in the literature, and again find large coefficient estimates. They conclude that the 1993 expansion "increased the extensive margin of labor supply separate and apart from any pro-work reforms to state welfare systems that occurred at the same time." In effect, the literature's consensus was and remains that the standard tests of the identifying assumptions (e.g. parallel pre-reform trends) and the use of judiciously chosen demographic and state-level covariates are adequate to address the potential bias from contemporaneous welfare changes. Moreover, there is not clear econometric guidance how to adjudicate such disagreements.

levels of educational attainment.)

[7]Appendix C extends this back-of-the-envelope calculation to illustrate how an equal proportionate reduction in welfare use–e.g. from removing non-working women from welfare use–produces large differences in employment levels and which produce spurious EITC estimates in DiD specifications.

3 Falsification Tests

Consider the following example of a "confounded" difference-in-difference estimator:

$$y_{it} = \beta_0 + \beta_1 \text{after}_t + \beta_2 \text{treat}_i + \beta_3(\text{treat}_i \times \text{after}_t) + X\beta + \gamma_s Z_s \times \text{after}_t + \varepsilon_{it}, \tag{2}$$

where the estimated model is the same as in equation (1) except that $\gamma_s Z_s$ captures a shock that affects some subgroup s only in the post-treatment period. Importantly, assume that s includes both treatment and control group subjects ($s \subseteq t, c$).

One interpretation of $\gamma_s Z_s$ is that it represents a variable whose intensity differs across treatment and control group (e.g. Z is the strength of the local economy, which varies between treated versus control states after the reform, and $\gamma_s = \gamma$). Another interpretation is that γ_s represents a post-reform shock to specific groups of individuals s, and Z_s is an indicator for each subgroup s, but those subgroups are not equally represented or balanced between the treatment and control group. Yet another potential interpretation is that $\gamma_s Z_s$ arises because of misspecification of the functional form of the estimator, for example, when when the time effect is logarithmic rather than linear, and subgroups s differ in their pre-reform levels of e.g. income.

In these cases, if the econometrician fails to control for $\gamma_s Z_s$ either directly, by rebalancing the sample, or modifying the functional form of the estimator, then this model will produce a biased estimate of β_3.

$\gamma_s Z_s$ may truly be an omitted variable that is simply not observed or measured in the data. However, there may be many potential groups s, and it may not be obvious or known which groups are subject to $\gamma_s Z_s$. When the potential number of subgroups s is large (as when there are many potential time-varying covariates), then it may be unfeasible to control for or balance on all, or to know which to choose.

Moreover, since $\gamma_s Z_s$ is only realized after the reform, it cannot be identified by examining pre-reform data. In particular, the analysis of "pre-trends" will not identify the source of the bias.

Finally, if one were to estimate $\gamma_s Z_s$ and find that it is correlated with treatment status and with the outcome, it may be ambiguous whether that is because it is truly an omitted variable or simply reflects heterogeneity in the treatment effect or even reflects the treatment effect itself, in which case including $\gamma_s Z_s$ would bias β in the opposite direction.

3.1 Placebo Tests

With this model in mind, I propose two tests to examine whether an estimator is confounded. To be clear, these tests cannot verify that an estimator is uncounfounded (which is why identifying assumptions are assumptions). They can, however, demonstrate whether an estimate is confounded and thus whether the identifying assumption is falsified.

First, I use placebo tests to assess whether the EITC DiD estimator produces a spurious treatment effect, and, if so, to diagnose the source of the spurious effect. As their name implies, placebo tests are used to test whether an estimator produces a treatment effect in a context constructed such that the actual treatment effect is zero. Assume the correct model, based on equation (2) is as follows:

$$y_{it} = \beta_0 + \beta_1 \text{after}_t + \beta_2 \text{treat}_i + \beta_3(\text{treat}_i \times \text{after}_t) + X\beta + \gamma Z \text{after}_t + \varepsilon_{it}$$

and assume that Z is correlated with treatment. Estimating equation (1) without including Z results in classic omitted variables bias:

$$\hat{\beta} = \beta_3 + \gamma \frac{\text{cov}(\text{treatment} \times \text{after}, Z)}{\text{var}(\text{treatment} \times \text{after})}$$

When the treatment $\times$ after effect is a dummy variable, then the bias term simplifies to $\gamma(\Delta Z)$, which is the effect of the omitted variable on y times the difference in the mean of the omitted variable between the two groups ΔZ.

$$\hat{\beta} = \beta_3 + \gamma(\Delta Z) \tag{2}$$

This suggests the following placebo procedure: estimate the DiD model in a context in which the true treatment effect is the same (i.e. when there is no treatment difference between groups) and thus β_3 is expected to be zero. If the estimates of $\hat{\beta}$ are non-zero (beyond that expected from sampling error), then that suggests bias. Moreover, to the extent that the estimated placebo effects $\hat{\beta}$ are correlated with characteristics of the sample that are omitted from the regression (ΔZ), that indicates that those characteristics are omitted variables.

While we usually think of omitted variables bias as arising when the variable is unmeasured or otherwise unavailable to the researcher (e.g. unmeasured ability when estimating the returns to schooling), the present case is unusual in that the potential confounding factors, such as exposure to welfare reform policies, are potentially measurable in the data and may simply be omitted by oversight or out of a belief that the other covariates (such as state by year effects and other demographic characteristics) are sufficient.

3.2 Estimating "Non-Parallel Trends" Bias

The previous section assumes that confoundedness arises from an omitted variable. Another cause of confoundedness in this context could arise because of imbalances in the composition of the treatment and control groups. Assume again the simplest 2x2 version of equation (2), where superscripts y, c index treatment and control group, and the subscripts 0,1 index time before and after treatment. Assume X is the same across groups and can be ignored, and the specification includes a common time effect captured by β_1.

In addition, assume that γ_s represents the magnitude of subgroup specific shock (a post-reform fixed effect), Z_s is an indicator for subgroup s, and ω_s^i represents the weight of each subgroup s in treatment group i. Note that because we have partialed out the common time effect (β_1), γ_s is simply the pre-versus post change in the outcome of each specific subgroup–its “trend” (Δy_s^i).

The DD estimator β_3 is given by:

$$\beta_3 = (y_1^t - y_0^t) - (y_1^c - y_0^c)$$

$$= \Delta y^t - \Delta y^c$$

Now decompose the mean outcomes of the treatment and control group into subgroups indexed by s:

$$= \sum \omega_s^t \Delta y_s^t - \sum \omega_s^c \Delta y_s^c$$

Adding and subtracting the change in the control group weighted by the treatment group weights ($\omega_s^t \Delta y_s^c$) (as in a basic Kitawaga-Oxaca-Blinder decomposition) and rearranging gives:

$$= \sum \omega_s^t (\Delta y_s^t - \Delta y_s^c) + \sum (\omega_s^t - \omega_s^c) \Delta y_s^c \tag{3}$$

The first component represents the weighted sum of the “treatment effects” calculated as the difference-in-difference in the outcome within each subgroup s, where the weights represent the fraction of the treatment group belonging to subgroup s.

The second component represents the quantity “explained” by differences in the composition of the treatment and control group holding fixed the coefficient estimates (the time trend or γ_s) estimated within the control group. Note that in this decomposition, this latter term contains no information on the post-reform outcome of the treatment group.

If the latter component in (3) is not zero, however, then even in the absence of any treatment effect, the counterfactual outcomes of the two groups would not be parallel. Intuitively, when treatment and control groups differ in their composition, the examination of whether trends of subgroups within group diverge instead of evolve in parallel can provide an indication of whether the “parallel trends” identifying assumption holds.

Specifically, under the identifying assumptions of the DiD estimator, this second component should always sum to zero, however subgroups are defined.[8] Given the decomposition, the component can be zero in two intuitive scenarios: if (1) the trend is the same within all subgroups of the control group (Δy^c_s is a constant for all subgroups s) or (2) the weights in the treatment and control group are identical (so that $(\omega^t_s - \omega^c_s)$ is zero for all s).[9]

Hence, this decomposition suggests an alternative falsification test for the identifying assumptions: form subgroups of the control group sharing similar characteristics (such as by demographic characteristics, geography, pre-reform propensity to work or use welfare, etc.), and then examine whether the weights of those subgroups differ between treatment and control group, and whether their outcomes experience divergent "trends." In short, are there ways to partition the sample such that the latter component of equation (3) is nonzero? If so, how much of the estimated treatment effect is "explained" by these divergent trends?

4 Analysis

4.1 Placebo Implementation

I assign placebo treatment and control group status to groups formed from individuals that were treated equally by the EITC (e.g. mothers of one child), and then re-estimate the key test statistic (the DiD estimate of the treatment effect) that estimate the "treatment" of one subgroup of mothers of one child to another "control" subgroup of mothers with one child. I estimate these placebo DiD models within treatment and control groups to deliberately exclude any comparison between differently-treated individuals.

In these placebo tests, the EITC treatment effect is the EITC indicator variable constructed following Strain and Schanzenbach (2021). In each specification I estimate the

[8]This is a relatively weak assumption—it might be realistic to make stronger assumptions on subgroups such that subgroup trends should be parallel between treatment and control groups.

[9]A third case, which I do not explore, arises if the imbalance in weights is inversely correlated with the subgroup trends and causes the sum of the weighted trends to be zero. While this does not imply a violation of the parallel trends assumption, it may imply that additional scrutiny of the research design may be warranted.

effect of the EITC treatment from a pairwise comparison between one randomly defined placebo treatment group and one placebo control group both drawn from the same actual treatment or control group. Following the structure of the policy (in which treatment assignment is based on a common demographic characteristic: number of children), I form placebo groups based on groups with similar pre-reform characteristics.

In practice, these placebo groups can be formed by arbitrary combinations of demographic characteristics. However, because the posited confounding variable is related to welfare reform, I form placebo groups based on demographic characteristics that predict similar pre-reform exposure to welfare use, nonemployment, or earnings. Specifically, I estimate the predicted likelihood of welfare use and nonemployment using a probit estimator on the same covariates described in the discussion of equation (1) including demographic characteristics (age, marital status), educational attainment, state of residence, and family composition (number of children and age of youngest child), and also a predicted log wage using the same characteristics in linear regression. I estimate these relationships exclusively in data from 1991, 1992, and 1993, before the major expansion of the EITC and the implementation of significant welfare reform changes.

I predict a pre-reform rate of welfare use, nonemployment, and wage for each individual in the sample, and form groups defined by deciles of each variable within each EITC treatment group (2+ children, one child, no children). Within each EITC treatment group, I assign one placebo group to be treatment and another to be control, estimate the DiD regression, and capture the coefficient estimate. I repeat this for all 90 potential permutations of the pairwise comparisons within each of the three EITC treatment groups. I repeat the exercise three times, one for each set of deciles formed by predicted welfare use, nonemployment, and earnings. (There are many ways to form these placebo groups, such as simply forming cells by arbitrary combinations of demographic characteristics; having tried many approaches the outcome of the analysis is the same.)

For each DiD estimator, I also measure the difference in predicted pre-reform welfare use between the placebo treatment and control groups, to assess whether the placebo coefficient estimates vary systematically with (omitted) characteristics of the two groups related to these pre-reform characteristics.

For comparison to these placebo estimates, I estimate actual DiD estimates based on the samples and comparisons in the four papers replicated above. Because these papers include more than two total treatment and control groups (e.g. mothers with 2+ children, one child, and childless women), I estimate all pairwise comparisons of treatment/control groups identified in these papers; the total EITC treatment effect is a weighted average of these pairwise comparisons (Goodman-Bacon (2021)). Collectively, these include comparisons across women with two or more, one, or zero children; all women and women with different levels of educational attainment (less than high school, high school, more than high school); and mothers with different-aged youngest children (age 0-2, 3-5, 6-12, 13+).

4.2 Results of the Placebo Tests

Figure 3 presents each placebo estimate as a function of the difference in the predicted pre-reform rates of welfare use between the treatment and control group. The figure only presents results from the permutations where the placebo treatment group's rate of welfare use exceeds that of the control group; the other permutations are simply a mirror image of this figure across the origin. First, Figure 3 shows that the placebo effects are often large and economically important—suggesting that "treated" groups experienced large increases in employment compared to the "control" group, even though these individuals are treated equally by the EITC.

Second, the magnitude of the placebo effects increases in line with the difference in the pre-reform rates of welfare use of the two groups being compared. Anytime a group with a higher rate of welfare use is compared with a group with lower rates of welfare use, the placebo estimates are biased toward finding larger treatment effects. However, when the two groups being compared have similar pre-reform rates of welfare use, the estimated treatment

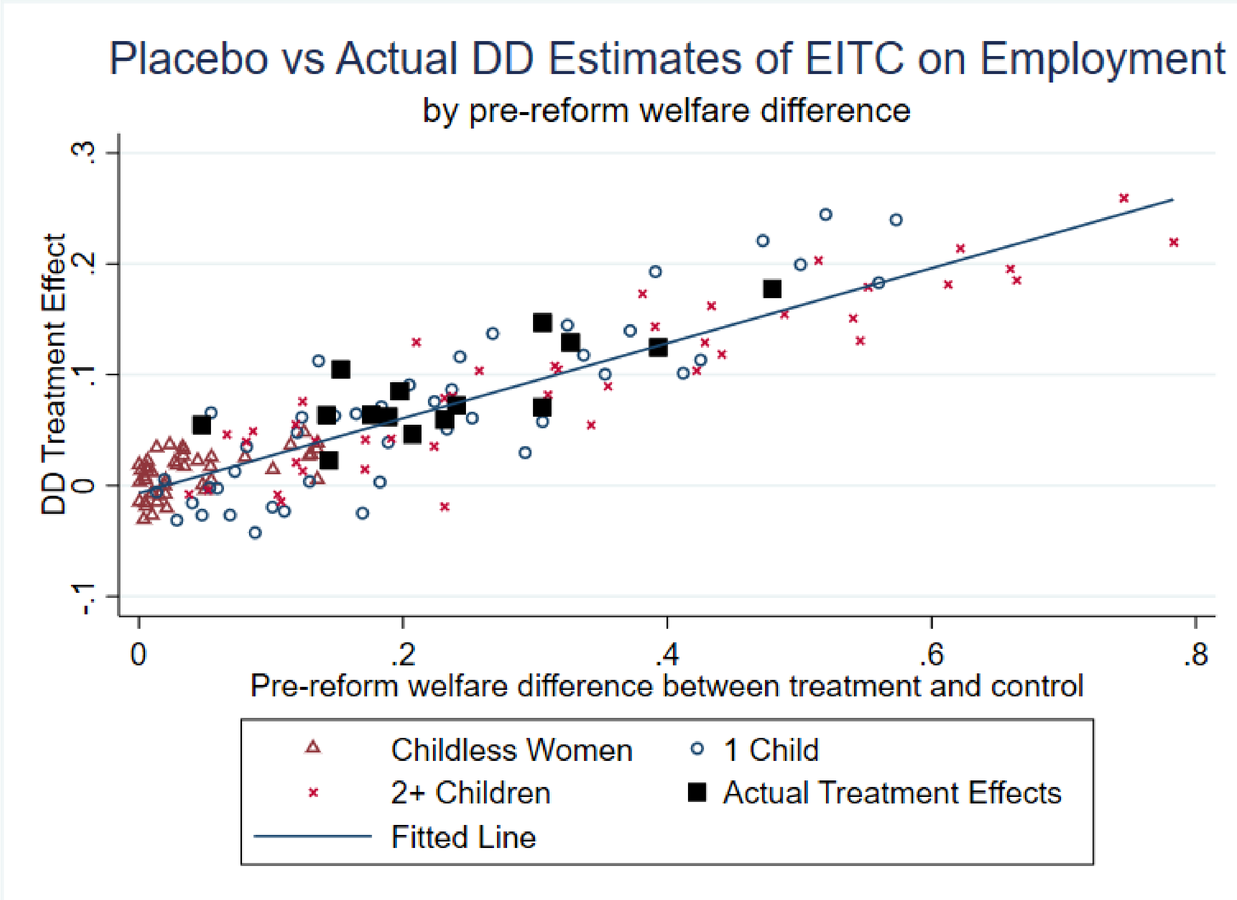

Figure 3: The figure presents placebo and actual DiD estimates arrayed by the gap in predicted pre-reform rates of welfare use between each treatment and control group for the permutations in which the treatment group's predicted welfare use exceeds that of the control group, and disaggregates the placebo effects based on whether they are estimated by pairwise comparisons of childless women, mothers with one child, and mothers of two or more children.

effect is close to zero, as indicated by the intercept of the fitted line.

Figure 3 also disaggregates the placebo effects into those from comparisons within childless women, mothers of one child, and mothers of 2+ children. The figure shows that within each group, the relationship between the placebo effect and the gap in welfare use is the same.

Finally, the black squares, which represent the actual DiD estimates (comparisons across groups whose actual EITC treatment was different) from the replicated literature, follow the same pattern. Knowing only the slope of the placebo-fitted line and the gap in pre-reform welfare use between actual treatment and control group, one can accurately predict what the actual DiD coefficient will be. In fact, in a regression of the actual DiD estimates on the values predicted by the fitted line in Figure 3, yields a coefficient of one and an intercept of zero.

In summary, the placebo tests suggest that the DiD estimator is indeed biased by a confounding effect that is proportional to pre-reform exposure to welfare use.

4.3 Non-Parallel Trends

To assess whether the bias term in (3) is likely to be zero, Figure 4 ranks single mothers by percentiles of pre-reform propensity to use welfare (the subgroups s) and separates the sample into the treatment and control groups (mothers with two or more children, mothers with one child, childless women). For each treatment group by subgroup s sample, the figure presents the percentage point change in employment between 1993 and 2000.

First, Figure 4 shows that the change in employment is strongly related to the predicted pre-reform rate of welfare use for each subgroup. Groups with low rates of pre-reform welfare use saw essentially no increase in employment, whereas groups with high rates of welfare use experienced increases in employment close to 40 percentage points. In other words, subgroups of the treatment and control groups are indeed "trending" at different rates.

Second, note that the treatment and control groups are not equally weighted in their composition. Mothers with two or more children are overrepresented in high-welfare use groups, and mothers with one child and childless women are overrepresented among groups with low pre-reform rates of welfare use. As a result, the bias term in (3) is not likely to be zero—distinct, identifiable subgroups are trending at different rates within group, and these subgroups are weighted differently across group, which means that even in the absence of treatment, there should be no expectation that the groups should trend in parallel.

How large is the resulting bias? Figure 4 suggests the change in employment is similar across treatment groups within percentiles of welfare use (subgroups). Among women with

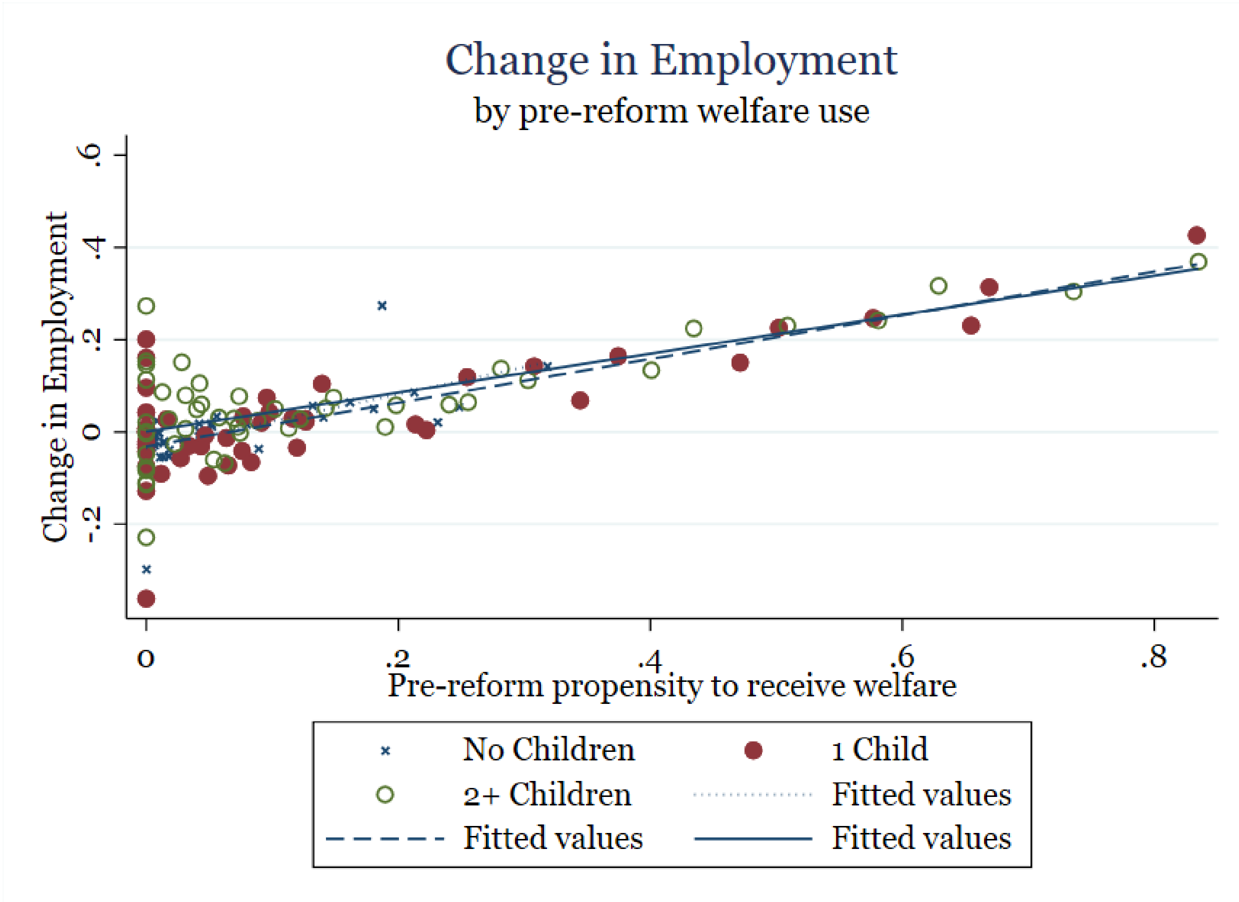

Figure 4: Average change in employment before vs after reform by pre-reform propensity to use welfare.

similar rates of pre-reform welfare use, the change in employment is similar whether they have two or more children, one child, or no children. Specifically, the fitted regression lines for each group are similar to each other, and their y-intercepts are close to zero, indicating that there is little difference in employment across groups after controlling for pre-reform rates of welfare use. In all, this analysis suggests the DiD estimate is primarily the result of bias rather than the treatment effect of the EITC. In other words, using the decomposition in (3), not only is the compositional bias term not zero, it appears to "explain" all of the DiD estimate.

To illustrate this, the top panel of Figure 5 starts with the standard graphical illustration of the DiD estimator comparing the employment rate of mothers of 2+ children and mothers

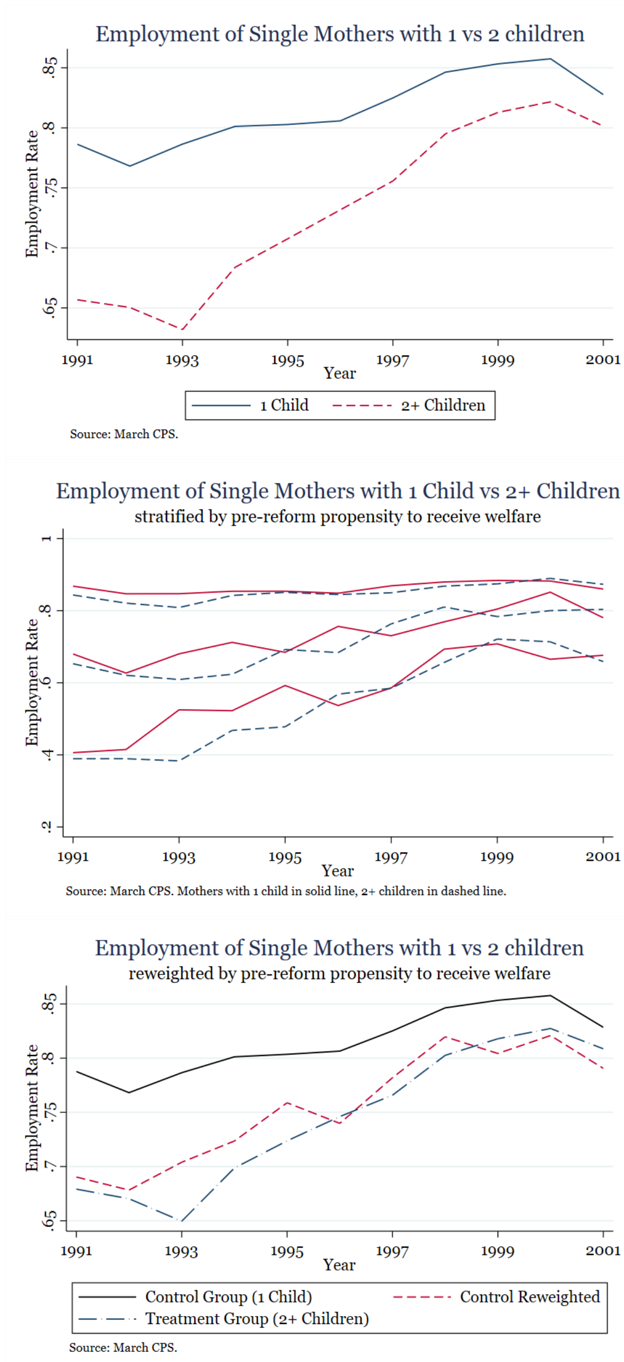

Figure 5: Note: Each panel compares the treatment group (blue) to the control group (red). The top panel provides the unweighted data. The second panel stratifies each sample by quintile of predicted pre-reform welfare use (as a proxy for exposure to welfare reform) and compares treatment- and control-group mothers with similar pre-reform rates of welfare use from the first, fourth, and fifth quintiles. The bottom panel re-weights the control group to have the same average predicted pre-reform exposure to welfare use as the treatment group.

with one child. The convergence in the employment rates of these groups after the 1993 expansion is the source of the estimated effect of the EITC in the DiD estimator. The middle panel of Figure 5 decomposes the treatment and control group into subgroups formed by quintiles of predicted pre-reform welfare use, similar to Figure 4. For clarity, the figure presents the annual employment rates of women in three of the five quintiles: the lowest predicted quintile of welfare use in 1993, the fourth highest, and the highest quintile.[10]

Figure 5 illustrates the considerable heterogeneity in subgroup trends within the groups. Whereas there is little change over time in employment of groups of mothers with low pre-reform rates of welfare use, employment surges among those with higher rates. Moreover, the time pattern of employment is extremely similar for individuals in the treatment and control group conditional on propensity to use welfare.

The bottom panel reweights the control group observations (represented unweighted as the black line) to match the pre-reform rates of welfare use observed in the treatment group (presented as the red dashed line), and compares the change in employment to that of the treatment group (the blue dashed line). The reweighted control group closely follows the employment rates of the treatment group. In fact, the reweighting of the control group more than explains the increase in employment over this period, suggesting that all of the difference-in-difference (the source of the variation in the DiD estimator) is associated with changes in employment of groups with different exposure to welfare.

5 Results Controlling for Confounding Effects of Welfare Reform

The analysis above suggests that exposure to welfare reform–where exposure is proxied, for example, by the predicted use of welfare in the early 1990s–is an omitted variable (or an

[10]The lowest three quintiles have similar rates of predicted welfare use and thus the second and third are omitted for graphical clarity.

important characteristic upon which to balance treatment and control groups). A direct way to address this bias is to include year-by-exposure effects as controls in the DiD specification. Figure 6 presents estimates of the effect of the EITC including controls for time effects

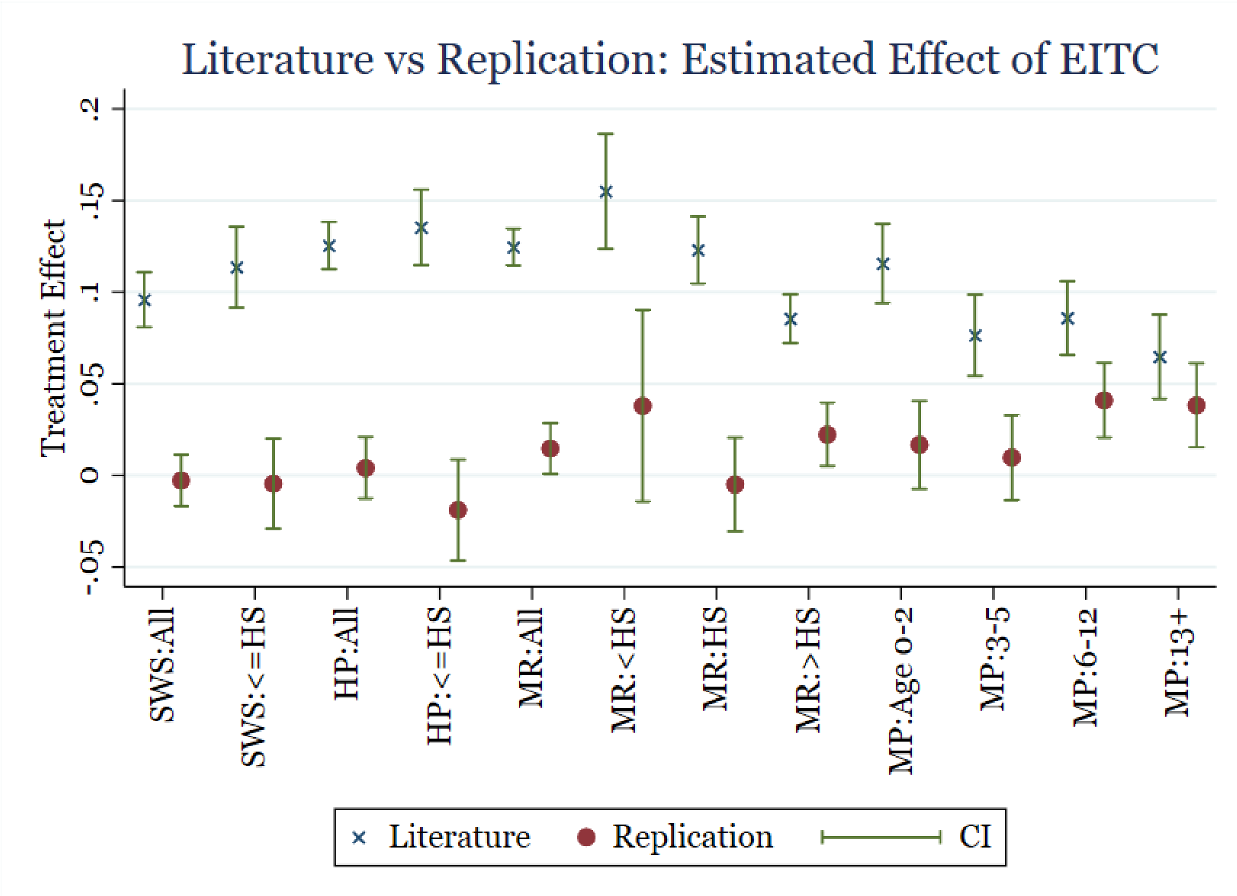

Figure 6: Note: This figure presents the estimated effects of the EITC in the same specifications presented in Figure 2 after controlling for year x pre-reform welfare use.

proportionate to exposure to welfare reform (as proxied by predicted pre-reform rates of welfare use). In contrast to the replication in Figure 2, the estimated effects are attenuated toward zero and typically not statistically significantly different from zero. In short, after controlling for these confounding factors, the EITC appears not to have increased the labor supply of single mothers.

5.1 Machine Learning Approaches to Covariate Selection

Some observers have been concerned that the estimated effect of the EITC may be confounded by contemporaneous changes in welfare policy. The analysis above suggests that the confounding variables were available in the data widely used in the literature. This raises the question of whether conventional approaches to identifying covariates to control for (or balance on) are effective. The choice of which covariates to include, or their functional form, is often ad hoc. Likewise, when researchers take care to re-balance treatment and control groups by matching or reweighting, it is not obvious which variables are important to balance.

Moreover, most approaches to matching exclusively use pre-reform information, which may exclude evidence of confounding effects. For instance, in synthetic control methods (e.g. Abadie and Gardeazabal 2003) or synthetic difference-in-difference methods (Arkhangelsky et al. 2021), treatment and control groups are reweighted to produce pre-reform trends that are parallel based on the relationship between time trends and characteristics in the pre-reform period. These methods deliberately exclude post-reform information, and as a result, they may exclude covariates that are unimportant in explaining time trends in the pre-reform period but are important in the post-reform period (confounders). The analysis above suggests that post-reform data can provide useful information regarding the validity of the empirical design.

Machine learning methods may provide a more systematic approach to identifying potentially confounding factors–covariates that are important predictors of post-reform outcomes. As the following analysis shows, they can be useful for generating hypotheses about potential confounders and for assessing the validity of the research design.

Figure 7 presents the actual employment rates of EITC treatment and control groups: single women and single mothers with one and two-plus children (solid lines). In addition, for each group, the figure presents the predicted out-of-sample machine learning employment rates of each group (dashed lines) from a simple machine learning (ML) algorithm (lasso) that selects covariates from a large set of demographic characteristics and pre-reform rates of welfare use interacted with year effects. Importantly, the predictions for each group are formed from out-of-sample predictions based exclusively on data from a different group. The predicted employment rates for mothers of two plus children and single women is estimated only from data on mothers of one child; predicted employment of mothers of one child is estimated from data on mothers of two children. As a result, the prediction contains no information on treatment status.

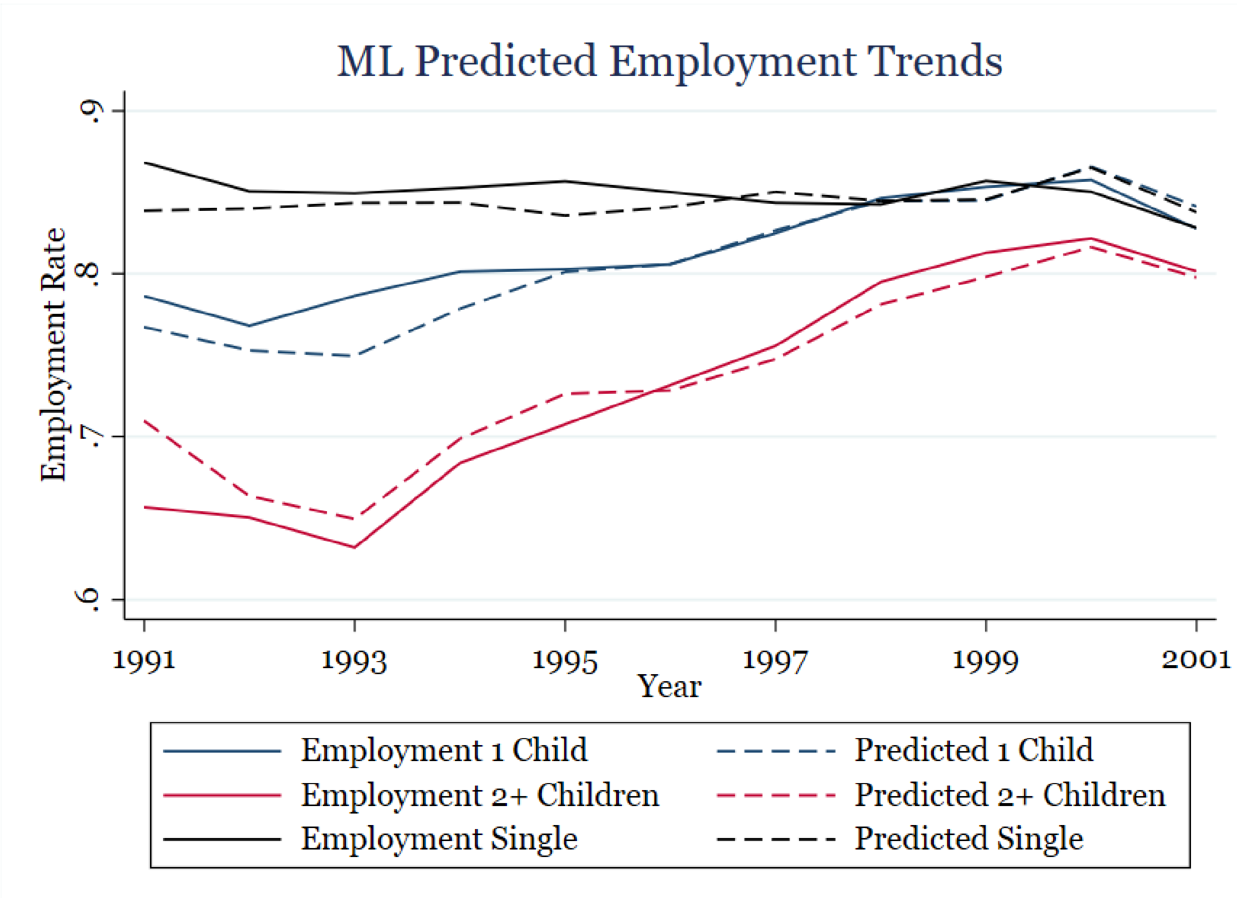

Figure 7: This figure presents actual employment rates of single women and single mothers with one and two plus children (solid lines) and the predicted out-of-sample machine learning employment rates of each group (dashed lines). The predicted employment rates for mothers of two plus children and single women is estimated exclusively from data on mothers of one child; predicted employment of mothers of one child is estimated from data on mothers of two children.

The analysis in Figure 7 suggests, first, that the out-of-sample predicted outcomes closely mirror the actual outcomes. However, this is evidence against finding a treatment effect. The out-of-sample prediction for mothers with two-plus children and childless women assumes they experience the same treatment effect as mothers with one child (and vice versa for the prediction for mothers of one child). That suggests that this representation of the counterfactual–the outcome of the treatment group (or control group) in the absence of treatment (or with treatment) is nearly identical to the actual outcomes of the groups. If the EITC had been an important cause of the differences in the changes in the employment rates of these groups, it should have caused large errors in the out-of-sample predicted outcomes–increases in employment in the control groups that were too large and increases in employment in the treatment groups that were too small.

Second, the ML analysis readily identifies time effects X age of youngest child and pre-reform welfare use as important predictors of the outcomes of individuals within each treatment group–the same confounding factors associated with exposure to welfare reform described above. With these covariates in hand, one could generate hypotheses about the appropriate specification to use and covariates to include in subsequent DiD analysis.

Finally, this analysis also allows a more general assessment of the likely validity of the research design, and whether the parallel trends is likely to have been true in the absence of treatment. In particular, the out-of-sample predicted outcomes suggest that the outcomes of treatment and control groups were unlikely to be parallel after the reform; indeed, if one only used data from one group (e.g. mothers of two or more children), one would expect a sharp convergence in the outcomes of all groups, partly because of within-subgroup trends in that group, and partly because the childless women group's employment rate is unlikely to increase, because it is already close to its maximum level.

6 Conclusion

The analysis in this paper suggests that the estimated effect of the 1993 Earned Income Tax Credit expansion is biased by confounding effects of welfare reform, and that the identifying assumption used to justify common DiD estimators are falsified in the data. One contribution of the paper is to propose methods to assess identifying assumption in such models; while identifying assumptions obviously cannot be proved to be true, they should at least not be contradicted in data available to researchers. These potential confounding effects are identified by placebo tests, examining within-group trends, and using ML methods.

In terms of economic outcomes, the analysis of this paper suggests that welfare reform policies had a larger impact on employment (and perhaps, other outcomes often attributed to the 1993 EITC expansion) than previously believed. Ethnographic narratives suggest that changes in the culture of welfare and the efforts to reduce caseloads were important contributors to changes in employment and welfare use; this paper suggests that exposure to such policies, as measured by predicted pre-reform rates of welfare use, explain a great deal of the change in employment during this time period.

While this paper exclusively analyzes the confounding effect of welfare reform on estimates of the effect of the 1993 EITC expansion on labor supply, the 1993 EITC expansion is widely used to analyze a variety of other outcomes using identical identification strategies. For example, the 1993 EITC is central to the evidence regarding the EITC's effects on poverty (Hoynes and Patel (2018)), childhood achievement (Dahl and Lochner (2012)), maternal health (Evans and Garthwaite (2014)), and infant health (Hoynes, Miller and Simon (2015)), among other outcomes. These studies may also be affected by the same bias, and may therefore attribute to the EITC changes caused by welfare reform policies, or caused by a combination of the mechanical increase in the after-tax income of households caused by the EITC and labor supply effects caused by welfare reform,

The evidence that the 1993 EITC expansion was responsible for the increases in the employment of single mothers has had long-lasting consequences for the design of social policy and the wellbeing of families. Based, in part, on the perceived efficacy of the 1993 EITC expansion, "virtually all gains in spending on the social safety net for children since 1990 have gone to families with earnings, and to families with income above the poverty line" according to Hoynes and Schanzenbach (2018). Relatively few resources are available to non-working parents, including those with very young children who were disproportionately affected by welfare reform. And recent efforts to increase transfers to non-working families remain contentious (see, e.g. Corinth et al. (2021)).

Moreover, the welfare consequences–the impact on the wellbeing of the single mothers affected by these policies–were more negative than widely believed. Rather than increasing employment because of the draw of tax subsidies for work, the increases in employment are closely associated with reductions in benefits for non-working mothers, which were intended to make those mothers worse off. Today, policymakers and advocates are debating new policies to provide cash benefits to parents which are not conditioned on work, such as universal basic income, refundable child tax credits, and paid parental leave. AFDC was, effectively, the country's largest paid leave program, and after its end, employment rates of single mothers of infant children increased more than 50 percent. One potential reason for today's interest in such policies may be lasting dissatisfaction with the post-welfare reform safety net, and its consequences for mothers with young children.

In evaluating the tradeoffs of these policies, the evidence in this paper suggests that the labor supply effects of limiting access to unconditional transfers in the 1990s were large. On the one hand, this suggests that such policies may reduce employment. On the other hand, this is evidence that the income effects are large, and thus that affected families have a high disutility of work, or benefit greatly from caring for young children, which may mean they are worth it.

References

Bastian, Jacob. 2020. "The Rise of Working Mothers and the 1975 Earned Income Tax Credit." *American Economic Journal: Economic Policy*, 12(3): 44–75.

Blank, Rebecca M. 2002. "Evaluating Welfare Reform in the United States." *Journal of Economic Literature*, 40(4): 1105–1166.

Corinth, Kevin, Bruce Meyer, Matthew Stadnicki, and Derek Wu. 2021. "The Anti-Poverty, Targeting, and Labor Supply Effects of the Proposed Child Tax Credit Expansion."

Dahl, Gordon B., and Lance Lochner. 2012. "The Impact of Family Income on Child Achievement: Evidence from the Earned Income Tax Credit." *American Economic Review*, 102(5): 1927–56.

DeParle, Jason. 2004. *American Dream: Three Women, Ten Kids, and a Nation's Drive to End Welfare.* Viking.

Eissa, Nada, and Jeffrey B. Liebman. 1996. "Labor Supply Response to the Earned Income Tax Credit." *The Quarterly Journal of Economics*, 111(2): 605–637.

Evans, William N., and Craig L. Garthwaite. 2014. "Giving Mom a Break: The Impact of Higher EITC Payments on Maternal Health." *American Economic Journal: Economic Policy*, 6(2): 258–90.

Flood, Sarah, Miriam King, Renae Rodgers, Steven Ruggles, J. Robert Warren, Daniel Backman, Annie Chen, Grace Cooper, Stephanie Richards, Megan Schouweiler, and Michael Westberry. 2024. "IPUMS CPS: Version 12.0 [dataset]."

Goodman-Bacon, Andrew. 2021. "Difference-in-differences with variation in treatment timing." *Journal of Econometrics*, 225(2): 254–277.

Grogger, Jeffrey. 2003. "The Effects of Time Limits, the EITC, and Other Policy Changes on Welfare Use, Work, and Income among Female-Headed Families." *The Review of Economics and Statistics*, 85(2): 394–408.

Grogger, Jeffrey. 2004. "Time Limits and Welfare Use." *The Journal of Human Resources*, 39(2): 405–424.

Hotz, V. Joseph, and John Karl Scholz. 2001. "The Earned Income Tax Credit." National Bureau of Economic Research Working Paper 8078.

Hoynes, Hilary, Doug Miller, and David Simon. 2015. "Income, the Earned Income Tax Credit, and Infant Health." *American Economic Journal: Economic Policy*, 7(1): 172–211.

Hoynes, Hilary W., and Ankur J. Patel. 2018. "Effective policy for reducing poverty and inequality? The Earned Income Tax Credit and the distribution of income." *Journal of Human Resources*, 53(4): 859–890.

Hoynes, Hilary W., and Diane Whitmore Schanzenbach. 2018. "Safety Net Investments in Children." *Brookings Papers on Economic Activity*, Spring: 89–150.

Kleven, Henrik. 2024. "The EITC and the Extensive Margin: A Reappraisal." *Journal of Public Economics*. Forthcoming.

Levine, Philip B., and Diane M. Whitmore. 1997. "The impact of welfare reform on the AFDC caseload." Vol. 90, 24–33, National Tax Association.

Looney, Adam, and Day Manoli. 2016. "Are there returns to experience at low-skill jobs? Evidence from single mothers in the United States over the 1990s."

Mead, Lawrence M. 2014. "Overselling the Earned Income Tax Credit." *National Affairs*, 61. Fall 2014.

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

Michelmore, Katherine, and Natasha Pilkauskas. 2021. "Tots and Teens: How Does Child's Age Influence Maternal Labor Supply and Child Care Response to the Earned Income Tax Credit?" *Journal of Labor Economics*, 39(4): 895–929.

Nichols, Austin, and Jesse Rothstein. 2015. "The Earned Income Tax Credit." *Economics of Means-Tested Transfer Programs in the United States, Volume 1*, 137–218. University of Chicago Press.

Strain, Michael R., and Diane Whitmore Schanzenbach. 2021. "Employment effects of the Earned Income Tax Credit: Taking the long view." *Tax Policy and the Economy*, 35(1): 87–129.

Weaver, R. Kent. 2000. "Front Matter." *Ending Welfare as We Know It*. Brookings Institution Press.

A Appendix

The summary statistics of the variables used in the analysis are presented in Table A.1.

Table A.1: Summary Statistics

	(1)	
	Mean	SD
Earned Any Income	0.82	0.38
Welfare Participation	0.10	0.29
Earnings	20897.82	25629.69
Age	31.97	8.86
No Children	0.62	0.49
One Child	0.17	0.37
Two or More Children	0.21	0.41
High School	0.30	0.46
Some College	0.35	0.48
College Graduate	0.22	0.41
Observations	150432	

Source: March CPS 1991-2001. Sample includes non-married women aged 20 to 50 (excluding childless women enrolled in school) and excludes single mothers whose youngest child is older than 18.

Summary statistics of the treatment variables used in the analysis, by number of children, are presented in Table A.2. The top panel are childless women, the middle panel is for mothers with one child, and the bottom panel corresponds to mothers with two or more children.

	1991	1992	1993	1994	1995	1996	1997	1998	1999	2000	2001
Work	0.87	0.85	0.85	0.85	0.86	0.85	0.84	0.84	0.86	0.85	0.83
Welfare	0.03	0.04	0.03	0.02	0.02	0.02	0.02	0.01	0.01	0.01	0.01
SWS Treatment	0	0	0	0	0	0	0	0	0	0	0
MR Tax if Work	-2,973	-2,935	-2,934	-2,901	-2,899	-2,907	-2,907	-2,907	-2,907	-2,907	-2,907
HP Sim EITC	0	0	0	40	40	40	40	41	40	39	40
Work	0.79	0.77	0.79	0.80	0.80	0.81	0.82	0.85	0.85	0.86	0.83
Welfare	0.23	0.24	0.24	0.22	0.19	0.18	0.15	0.12	0.10	0.08	0.07
SWS Treatment	0	0	0	0.92	1	1	1	1	1	1	1
MR Tax if Work	-1083	-1001	-955	-702	-633	-609	-609	-609	-609	-609	-609
HP Sim EITC	625	693	729	957	1,010	1,013	924	948	941	917	927
MP Average EITC	711	781	825	1,092	1,155	1,160	1,172	1,204	1,193	1,166	1,185
Work	0.66	0.65	0.63	0.68	0.71	0.73	0.76	0.80	0.81	0.82	0.80
Welfare	0.38	0.37	0.39	0.35	0.31	0.29	0.25	0.20	0.16	0.14	0.11
SWS Treatment	0	0	0	0.5	0.78	1	1	1	1	1	1
MR Tax if Work	-651	-554	-501	-19	291	548	548	548	548	548	548
HP Sim EITC	633	707	745	1,194	1,457	1,670	1,440	1,473	1,469	1,436	1,452
MP Average EITC	780	865	916	1,453	1,769	2,014	2,028	2,079	2,060	2,024	2,048

B Time Limits

In a related literature, Grogger (Grogger (2003) Grogger (2004)) finds that welfare reform's newly required five-year time limit on cumulative welfare use was a significant contributor to the decline in welfare participation and increase in employment during the late 1990s.

This literature is relevant to the concern that the EITC is confounded by the effects of welfare reform because the evidence on the effects of welfare time limits uses a nearly identical identification strategy and regression specification, but very different treatment and control groups based on the age of children (but not the number of children). As a result, if welfare reform is a confounding factor in the analysis of the EITC it is also likely to have the same counfounding effect on estimates of welfare time limits. Moreover, finding a counfounding effect of welfare reform on time limits corroborates and reinforces the analysis above.

In theory, these time limits generate an immediate incentive to reduce welfare use to conserve benefits as insurance against future economic shocks. However, this precautionary motive should only affect families with younger children because the typical five-year time limit is not binding on families where the youngest child is aged 13 to 17; their eligibility already will end when the child turns 18 before the limit could be reached. Grogger shows that the level of welfare participation of families with young children fell faster than that of families with older children after the implementation of time limits and interprets this as evidence of an anticipatory response to time limits.

The econometric analysis of the effect of welfare time limits follows the same DiD specification as the EITC with the addition of an additional set of policy variables for the impact of time limits. In particular, in Grogger's 2004 specification, the effect of time limits is captured with a dummy variable for whether the state has a welfare time limit in place in that year interacted with two age-of-youngest child dummy variables (for "less than age 6" and "age 6 to 12") plus dummy variables the age of youngest child.

B.1 Analysis

In my reanalysis of time limits, I replicate the same general approach as I take to the EITC. First I re-produce the analysis in a consistent sample from 1991-2001 and compare it to published estimates, corroborating the original analysis. In my replication, I exclude mothers whose oldest child is 18, include the maximum value of the EITC in each year for each group (single women, mothers with one child, mothers of 2+ children). This approach combines the simplest specifications from Grogger 2003 and 2004 rather than replicating each of multiple analyses for clarity and simplicity, though the results of the analysis are the same applying the exact specifications.

Second, I compare the DiD coefficient estimates to placebo estimates and show that these estimates are similarly predictable based on the ex-ante differences in welfare use between treatment and control observations.

Finally, I include controls for ex-ante welfare use by time effects and show the estimated effect of time limits is attenuated.

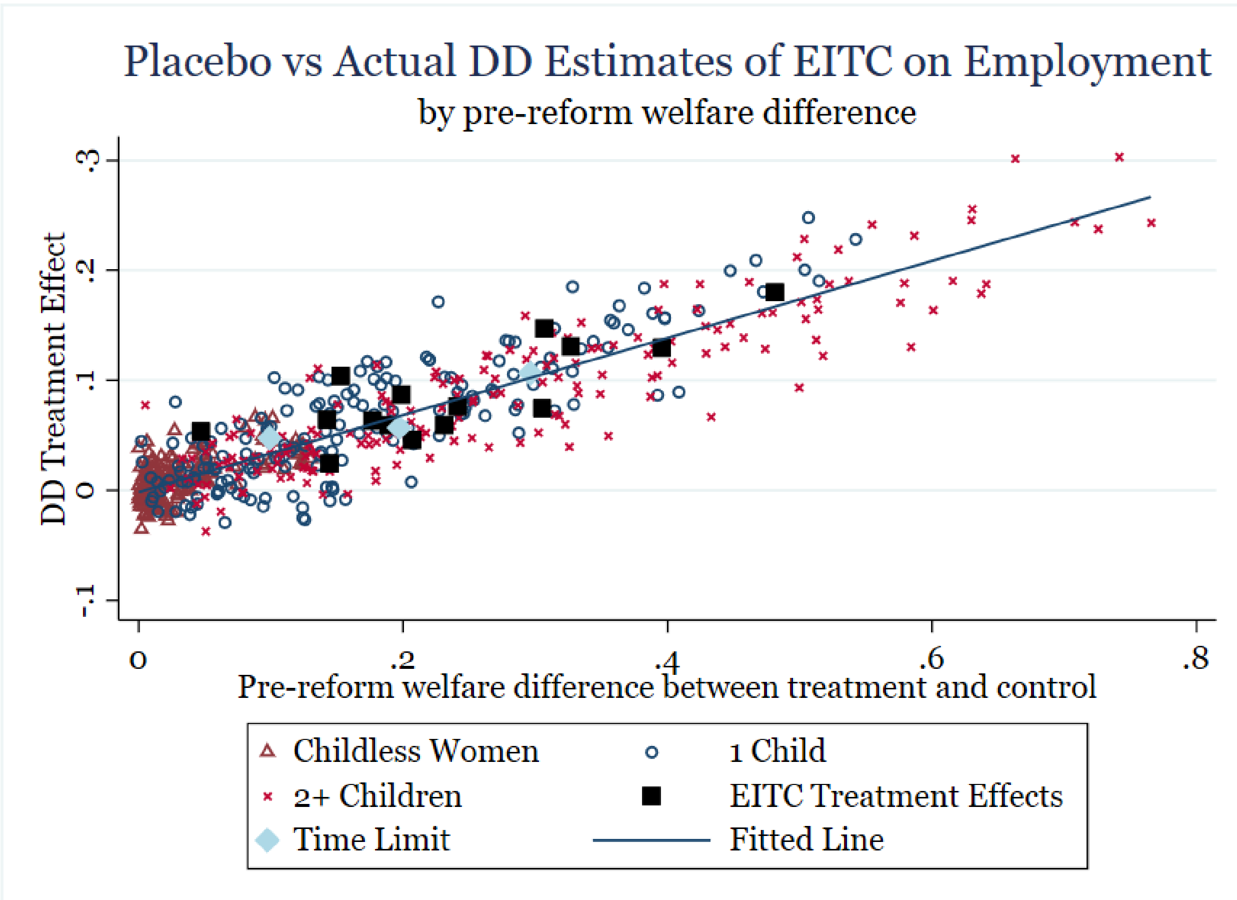

Figure A.1: Plots placebo treatment effects and actual estimated treatment effects of the EITC (black squares) and welfare time limits (blue diamonds).

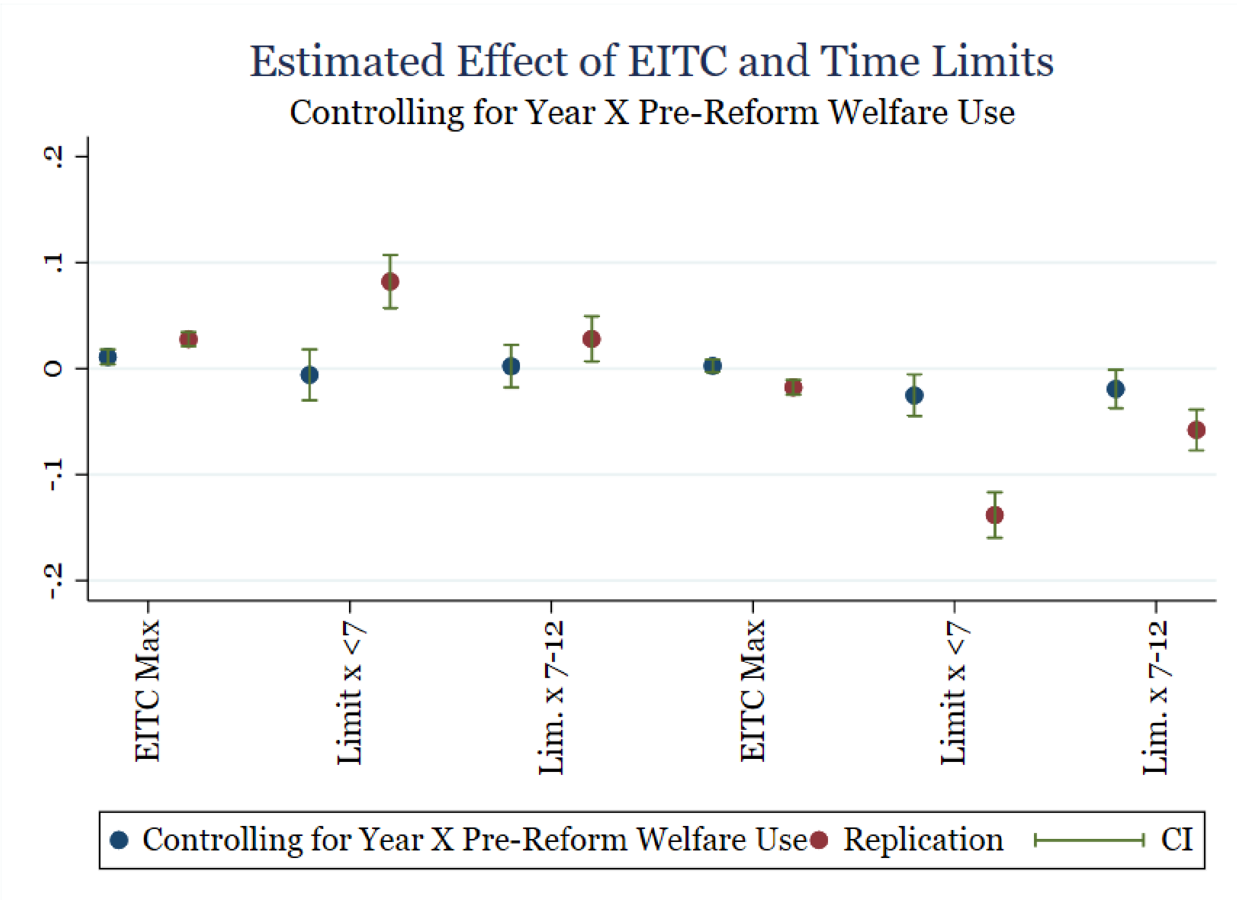

Figure A.2: Compares estimates of the effects of the EITC and welfare time limits on welfare use with and without controls for year-by-pre-reform welfare use.

C Back-of-the-Envelope Simulation

The caseload reduction hypothesis makes specific empirical predictions about which mothers would be affected and by how much. If the policy operated by reducing the likelihood of successfully applying for or renewing welfare benefits and was implemented at the caseworker-recipient level, its effects should be proportional to the rate of welfare use before the reform and the gap between the employment of single mothers and comparable single women. However, because there were large differenes in the propensity to use welfare in the early 1990s, an equal proportional decline in welfare use (and increase in employment) would result in large differences in levels (measured, for example, as percentage-point changes).

Before reform, for example, 32 percent of single mothers reported receiving cash welfare. After reform, the rate was 11 percent, a 65 percent (or 21 percentage point) decline. The decline in welfare among mothers of one child was 16 percentage points (from 24 percent to 8 percent), while the decline among mothers with two or more children was 24 percentage points (from 38 percent to 14 percent). Likewise, the rate of welfare use of mothers with children less than 3 fell by 34 percentage points, whereas it fell by only 6 percentage points among mothers whose youngest child was 13 or older.

C.1 Simulation

To illustrate, I simulate a policy intervention that removes single mothers from welfare at random and in numbers sufficient to achieve annual caseload reductions observed in the March CPS. Nonworking mothers are assigned to find work at random. By design, the "effect" of the simulated policy is unconditional on the characteristics of the mother (such as level of education, number of children, or age of youngest child). Implemented in the March CPS, I compare the effect of this placebo policy, both qualitatively and econometrically, to the evidence of the EITC's effect, including the difference-in-difference estimators widely used in this literature.

The simulation procedure is as follows: I construct a synthetic panel data set of annual micro data from the March CPS. I start with a pre-reform March CPS cross section (e.g. data from 1993) of single mothers and childless women (identical to that used in the analysis above), stack ten of those cross sections together, and label each cross section with a different year from 1991 to 2001. Then, for each of those "synthetic" years, I randomly assign a fraction of welfare recipients to be "kicked off" in proportion to the amount needed to achieve the actual fraction of individuals on welfare in each year from 1991 through 2001. For example, in 1993 33 percent of single mothers received welfare in the March CPS. In 2000, only 10 percent received welfare, a 68 percent reduction. Hence, for the synthetic year 2000 cross section, I randomly assign 68 percent of observed welfare recipients to not be on welfare (i.e. the dummy variable is recoded from 1 to 0 in 68 percent of cases). Likewise, if the employment rate rises from 70 percent to 80 percent between 1993 and a subsequent year, then I assume the job finding was at random: I assign 33 percent of nonworking mothers ((70-80)/(100-70)) to be employed. The result of the exercise is a 10-year panel dataset with the same covariates as are available in the March CPS, but in which the only time series variation is the result of randomly assigning observations to not be on welfare and work.

I estimate the same regression specifications as above. The results are presented in the following figure and compared to my replication of each specification. The coefficient estimates from the simulated data are similar to those of the actual analysis. The implication is that an equal proportional change to welfare use–such as changes in bureaucratic procedures and imposition of work requirements–results in identical coefficient estimates when estimated in levels as observed in the literature. Beyond reproducing the same DiD coefficients, this simulation also rationalizes other "puzzles" in the data, like increases in employment among mothers with more than two or three children, and by the age of a mother's youngest child (not shown).

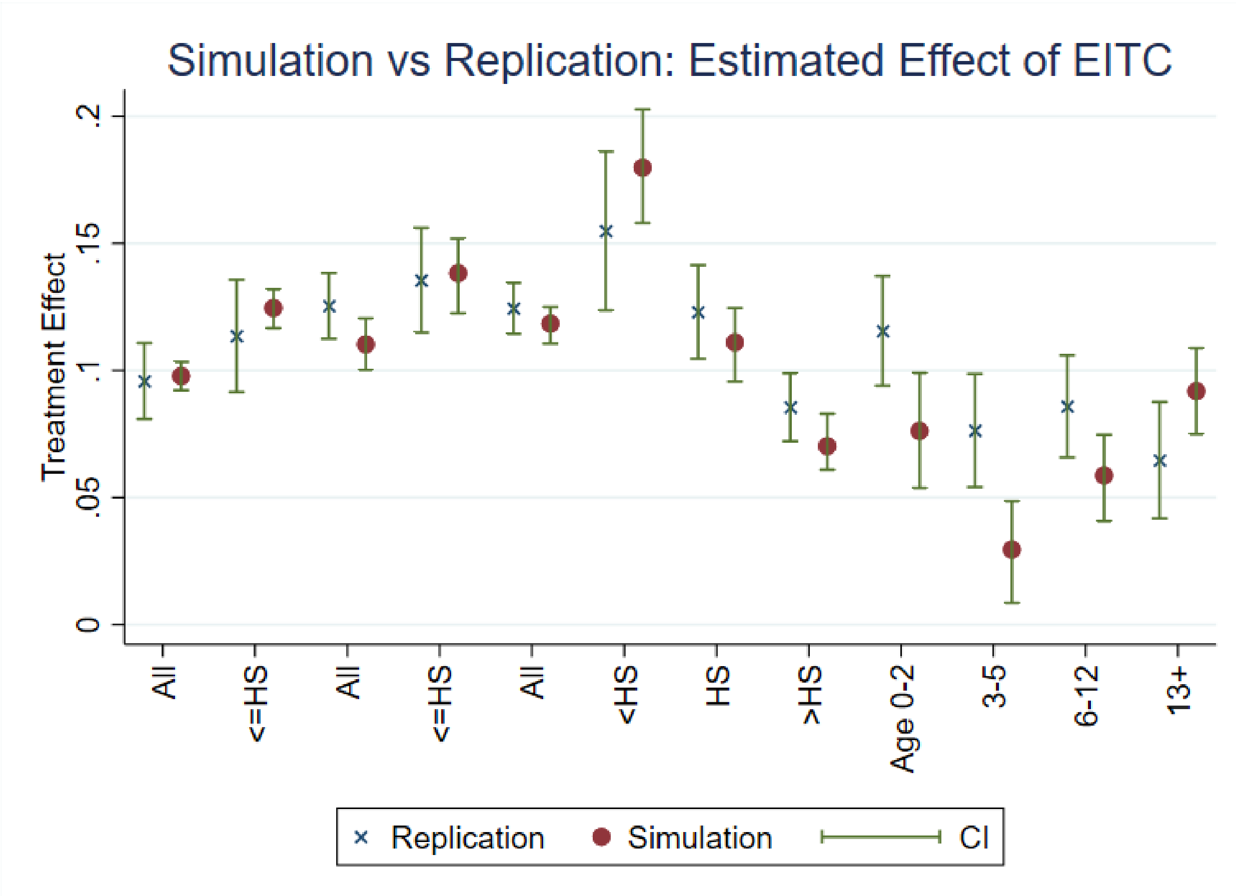

Figure A.3: Note: Compares coefficients on EITC treatment variables in replications of DiD estimators in March CPS to coefficients estimated on simulated data where the only time series variation is caused by equal proportional increases in employment at random across groups.