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

Adam Looney†

January 2026

Abstract

A large literature credits the 1993 Earned Income Tax Credit (EITC) expansion with increasing labor supply among single mothers in the 1990s, using difference-in-differences designs that compare mothers with different numbers of children. I show that, once contemporaneous welfare reforms are properly accounted for, the estimated effect of the EITC falls to zero in replications of four representative studies. The original designs violate the parallel trends assumption because treatment and control groups differ systematically in exposure to welfare policy changes. Placebo tests and a reweighting exercise demonstrate that the canonical estimates arise entirely from compositional imbalance between treatment and control groups. A simple machine-learning procedure would have flagged this violation ex ante. The results imply that welfare reform, rather than the EITC expansion, drove the employment gains of the 1990s.

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

*I am grateful to seminar participants at the University of Chicago, the National Tax Association, IIPF, NBER, and the Brookings Institution, and to numerous colleagues for helpful comments and conversations. 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

A large body of research credits the 1993 expansion of the Earned Income Tax Credit (EITC) with the large and persistent increase in labor supply and the decline in welfare use among single mothers in the 1990s. In fact, the 1993 expansion is particularly important because it underpins much of what we know about the EITC's labor supply effects [Nichols and Rothstein, 2016] and is also the basis for estimates of its effects on broader outcomes—including earnings, income, poverty, health, and human capital. In part because the 1990s changes are viewed as driven by the EITC's work-contingent subsidies, they contributed to a shift in social policy after which "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 1990s experience continues to shape today's debates over child-related tax and transfer policy.

A longstanding concern is that the estimated effects of the 1993 EITC are confounded by the contemporaneous changes to cash welfare—"welfare reform"—implemented through federal law and earlier state waivers [e.g., Mead, 2014, Kleven, 2024]. In particular, ethnographic histories suggest that administrative barriers and changing norms associated with welfare reform were central drivers of declining welfare caseloads and rising employment [DeParle, 2004, Weaver, 2000]. Standard difference-in-differences (DiD) estimates of the effect of the EITC address this concern by including state-by-year fixed effects or similar controls for state-level policies and economic conditions. The concern, however, is that welfare reform affected individuals in proportion to their reliance on the welfare system. If so, the relevant variation is within states rather than across them, and state-by-year controls are insufficient to eliminate bias.

In this paper, I show that the identifying assumptions underlying canonical DiD designs are violated. Treatment and control groups—mothers with different numbers of children—differ systematically in their exposure to welfare reform, and this compositional imbalance drives the estimated effects. When DiD specifications are augmented to include controls for the interaction of year dummies with pre-reform welfare exposure—time shocks proportionate to welfare use—the estimated effect of the EITC is attenuated to zero in all of the studies, and these time effects explain *all* of the increase in employment in the 1990s.

While this suggests that welfare reform is a source of bias, a natural concern is that controlling for welfare exposure absorbs genuine EITC variation, since EITC treatment and welfare exposure are correlated. To formalize the confounding mechanism and address this concern, I propose two tests. First, I re-draw the standard time series graphs illustrating the differences between EITC treatment and control groups over time after reweighting the sample to equalize predicted pre-reform welfare exposure between groups. This shows that the relative changes in employment that identify the DiD estimate of the EITC arise entirely from differences in the composition of the treatment and control groups. Second, I estimate "placebo" DiD regressions that compare equally-EITC-treated individuals, and show these specifications generate large, spurious "effects" that are directly proportional to the difference in predicted pre-reform welfare use between compared groups. I show these placebo effects are evidence of omitted variables bias, and the source of the bias is the exclusion of controls for exposure to welfare reform.

From the perspective of empirical practice, it is undesirable that this source of bias remained undetected for so long, despite it being available in the data. To address this problem prospectively, I propose using a simple machine-learning (ML) algorithm for covariate selection applied to the standard graphical analysis used to assess the "parallel trends" assumption in DiD estimators. Specifically, I ask the ML model to select covariates that predict annual employment within one group (e.g., the control group) and use those covariates to out-of-sample forecast outcomes in the other group (e.g., the treatment group). This provides a prediction of the counterfactual outcome that is constructed to exclude the effect of the difference in treatment. Using the example of the EITC, this procedure selects time effects in pre-reform welfare use as important covariates, and the resulting predicted counterfactual outcomes closely match realized outcomes, leaving no room for an EITC-driven treatment effect. In short, this visualization would have caused researchers to reject the parallel trends assumption ex ante.

This paper makes three contributions. First, it shows that the canonical difference-in-differences designs used to evaluate the 1993 EITC expansion are not identified. Reweighting and placebo tests demonstrate that the widely cited employment effects are entirely attributable to this compositional imbalance, not to differential EITC exposure. Second, it provides a simple corrective specification. Augmenting four influential studies with a single control—Year $\times$ predicted pre-reform welfare exposure—reduces the estimated EITC effects to zero in every case. The results imply that welfare reform, rather than the EITC, explains the employment gains of the 1990s. Third, it contributes a general design diagnostic. A simple machine-learning cross-prediction procedure provides a transparent graphical test of parallel trends. This approach can be applied broadly in difference-in-differences settings to detect violations before causal claims are made.

The rest of the paper proceeds as follows. Section 2 provides background and introduces the proportional-exposure hypothesis, with descriptive evidence motivating the reanalysis. Section 3 reports replications of canonical EITC specifications, showing that controlling for time $\times$ baseline welfare exposure eliminates the estimated effects, and links to replication details in Appendix B. Section 4 explains why the bias arises, using a decomposition of the DiD estimator, reweighting, and placebo tests that demonstrate the spurious nature of the canonical estimates. Section 5 uses machine-learning cross-predictions as a design diagnostic to surface the omitted confounder ex ante. Section 6 concludes with implications for research that relies on the 1993 expansion (poverty, child outcomes, maternal and infant health) and for current policy debates.

2 Background and Literature Review

An extensive literature credits the 1993 Earned Income Tax Credit (EITC) expansion with historic increases in employment and reductions in welfare receipt among single mothers in the 1990s [see reviews in Blank, 2002, Hotz and Scholz, 2001, Nichols and Rothstein, 2016]. These conclusions are based on difference-in-differences (DiD) designs comparing more-treated groups (e.g., mothers of two or more children) to less-treated or untreated groups (mothers of one child or single women without children), and attributing the post-1993 divergence in employment trends to the EITC.

The 1993 expansion was the largest in the program's history: the maximum credit for families with two or more children nearly doubled from $1,511 to $3,556 (1993 dollars), the phase-in rate rose from 19.5 to 40 percent, and a small credit was extended to childless workers for the first time. A typical estimating equation has the following form:

$$y_{ist} = \theta_{st} + \beta\,\mathrm{EITC}_{ist} + \gamma\,\mathrm{Children}_i + X_{ist}\delta + \epsilon_{ist}, \tag{1}$$

where y_{ist} is an indicator for employment of individual i in state s and year t, and θ_{st} are state $\times$ year fixed effects. EITC_{ist} captures variation in EITC generosity, and $\mathrm{Children}_i$ are dummies for the number of children. The vector X_{ist} includes standard demographic controls such as education, race, marital status, age, and age of youngest child. The identifying assumption of this model is that, absent changes in EITC policy, outcomes for individuals with different numbers of children within a given state would have followed parallel trends over time, conditional on state $\times$ year effects and the covariates X_{ist}. However, the mid-1990s also saw sweeping welfare reforms—both through federal legislation and earlier state "waivers" allowing states to implement pro-work changes in program rules and sanctions for non-compliance—that dramatically reduced caseloads. Ethnographic accounts suggest that these changes and associated administrative barriers to welfare receipt were the major causes of declining welfare use and rising employment 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 [welfare] caseloads" (as quoted in Weaver [2000]). In areas with the most dramatic reductions in welfare use, like Wisconsin, ethnographers attributed the changes to work requirements enforced by new bureaucratic procedures (DeParle [2004]). As one observer summarized, "the hassle factor of welfare is much more powerful in pushing them off the rolls (and consequently into jobs) than the vaguer promise of later wage subsidies" [as quoted in Mead, 2014].

The key aspect of these welfare policy changes was that they affected individuals directly in proportion to their exposure to the welfare system. Groups with higher baseline welfare use—such as mothers of two or more children—were more exposed to changes in the attitudes of their caseworkers or to being sanctioned under new rules, and thus may have experienced greater increases in employment regardless of EITC eligibility.

Figure 1 illustrates this concern. It presents a binned scatterplot of the change in employment of single mothers between the pre-reform (1991–1993) and post-reform (1999–2001) periods, plotted against each group's predicted pre-reform welfare exposure.[1]

The figure shows a strong positive relationship: groups with higher predicted welfare exposure experienced larger subsequent increases in employment. The slope of the regression line is 0.65, implying that a ten-percentage-point increase in predicted pre-reform welfare exposure for any group is associated with a 6.5 percentage-point increase in employment.

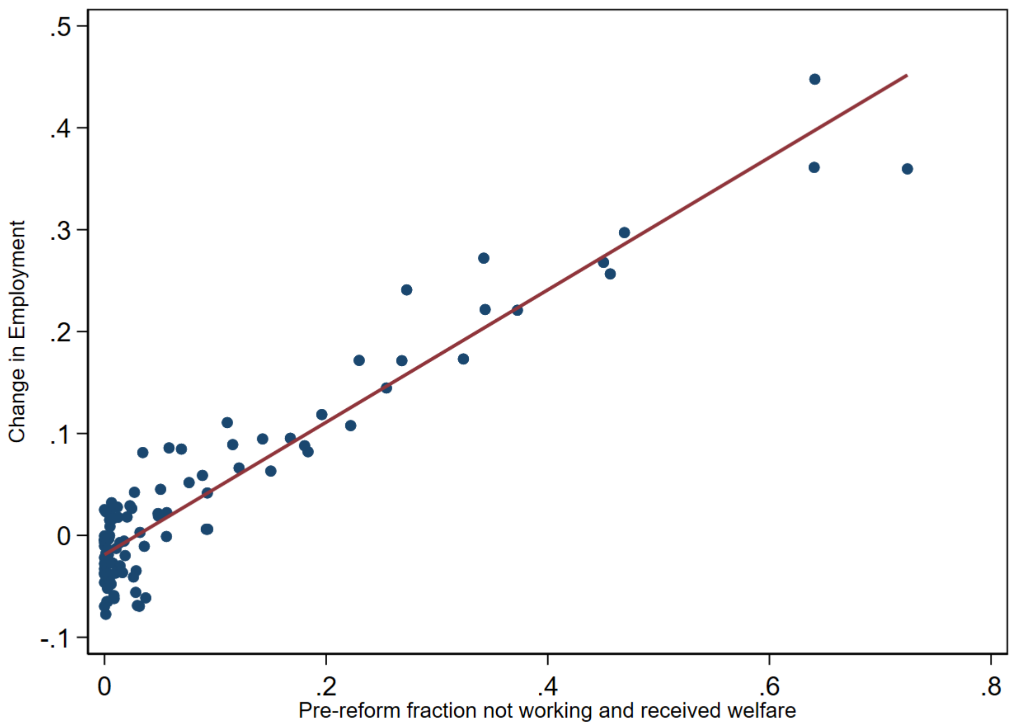
A back-of-the-envelope calculation illustrates the implications. In the pre-reform period, about 12.8 percent of single mothers with one child were on welfare and not working, compared to 26.4 percent of mothers with two or more children. Given the slope in Figure 1, this difference in baseline exposure predicts employment gains of roughly 8 and 17 percentage points, respectively. The actual changes were 7 and 17 percentage points. In other words, the canonical DiD estimate is fully predicted by baseline welfare exposure alone—without invoking any EITC effect.

Appendix E expands this calculation into a Monte Carlo exercise. Starting from a single

[1] Welfare exposure for individual i is defined as the predicted probability of being on welfare and not working in 1993: $\Pr(wno_i = 1 \mid X_i) = \Phi(X_i\beta)$, where X_i includes indicators for education, race, marital status, state of residence, age of youngest child, and a flexible polynomial in age with interactions by education. The model is estimated separately by number of children $(0, 1, 2+)$ using data from 1991 to 1993.

pre-reform cross section, I simulate a policy that randomly removes mothers from welfare and assigns non-working mothers to employment at rates matching the actual annual changes observed in the 1990s, with no differential treatment by number of children, education, or any other characteristic. This simple counterfactual policy reproduces the canonical DiD estimates almost exactly and replicates "puzzles" like the fanning-out of employment changes by family size and age of youngest child highlighted by Kleven [2024].

Figure 1: Employment Changes and Pre-Reform Welfare Exposure



Notes: Each point represents a percentile bin of single women grouped by predicted pre-reform probability of receiving welfare and not working. The horizontal axis shows the 1993 mean of predicted welfare exposure within each bin; the vertical axis shows the change in employment rate between 1993 and 2000. The fitted line has slope 0.65 (s.e. 0.02) and intercept −0.02. Sample: Single women ages 20–50 from March CPS 1991–1993 and 1999–2001 (Flood et al. [2024]).

Although Kleven [2024] raises the same underlying concern—arguing that the canonical evidence on the 1993 EITC expansion is confounded by contemporaneous welfare reform and emphasizing patterns such as the fanning-out of employment gains across family size—his analysis does not quantify the magnitude of bias in standard DiD estimates or provide an implementable correction. The analysis here formalizes that concern, measures the bias directly, and offers a specification that addresses it.

The absence of a quantitative resolution has allowed disagreement in the literature to persist. For example, Strain and Schanzenbach [2021] reject Kleven's critique, arguing that the additional controls in his empirical strategy absorb the true effect of the EITC. With specific regard to the 1993 expansion, they emphasize that state AFDC waivers created heterogeneity in the timing and intensity of welfare reform, and contend that the EITC expansion increased employment even in states that had not yet implemented waiver policies. On this basis, they conclude that the expansion raised employment independently of welfare reform and that the canonical interpretation of the 1990s evidence remains valid.[2]

Likewise, subsequent published work (e.g., Michelmore and Pilkauskas 2021) continues to rely on the same canonical DiD design without directly addressing this identification issue.

The central question, therefore, is whether the canonical difference-in-differences estimator is biased. Despite the prominence of this debate, the literature offers little econometric guidance on how to adjudicate it. The next sections propose and implement several empirical tests that establish the presence of bias, quantify its magnitude, and provide a corrective specification.

3 Replication Controlling for Welfare Exposure

I begin by showing that directly controlling for pre-reform welfare exposure eliminates the estimated effect of the EITC in the specifications used by four influential papers that examine the 1993 expansion: Strain and Schanzenbach [2021], Meyer and Rosenbaum [2001], Hoynes and Patel [2018], and Michelmore and Pilkauskas [2021]. These studies are representative of the published evidence on the EITC, use widely known data from the March CPS, but

[2]Figure 7 (Appendix D) shows that the relationship between baseline welfare exposure and subsequent employment growth is virtually identical in waiver and non-waiver states. This suggests that the proportional-exposure pattern is not driven by the timing of formal AFDC waivers, but reflects broader, nationwide changes associated with welfare reform.

differ in their parameterization of the EITC, the subpopulations studied, and whether they incorporate only federal or both federal and state expansions.

I focus on the period 1991–2001, which brackets the 1993 expansion and the 1996 federal welfare reform while excluding earlier and later policy changes.[3] Following the literature, I restrict the sample to non-married women aged 20–50 whose youngest child is under 18, and I include covariates standard in the literature: a quartic in age, dummies for education, race, age of youngest child, and marital status. All specifications include state-by-year fixed effects, and the basic estimating framework follows Equation 1.[4] The dependent variable is an indicator for employment last year (positive earnings), although similar results hold for alternative employment and earnings measures.

My replication (documented in Appendix B) closely mirrors the published results, and each specification produces economically large and statistically significant estimates of the 1993 EITC expansion's effect on employment.

Augmenting with pre-reform welfare exposure. I next show what happens when each specification is augmented with an interaction of year dummies and the *individual's pre-reform* welfare exposure rate.[5] This control captures the possibility that welfare reform effects were proportional to prior welfare use and therefore differentially affected treatment and control groups.

The results are shown in Figure 2. In every case, including this interaction term causes the EITC effect estimates to collapse toward zero and become statistically insignificant. This pattern holds across all four specifications, despite differences in EITC parameterization, sample restrictions, and treatment variables.

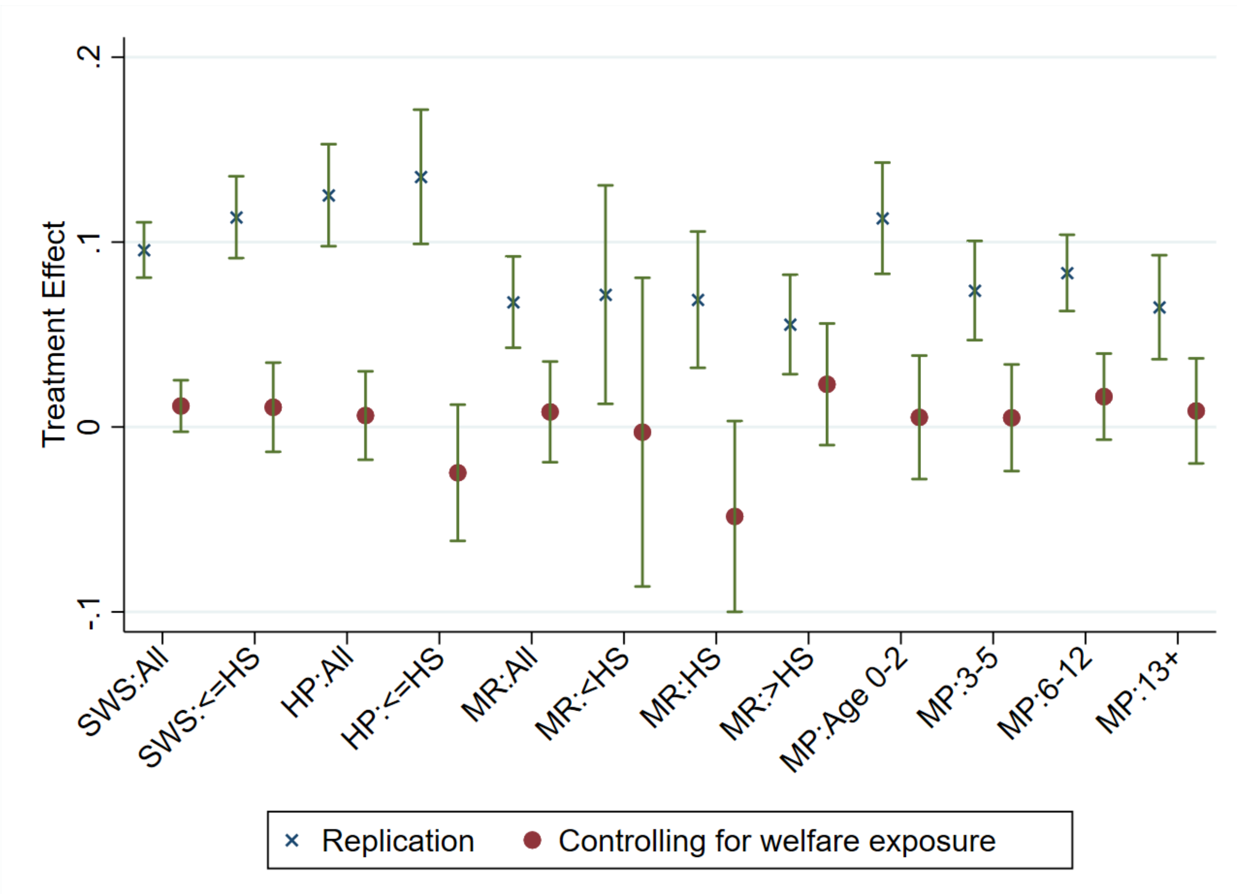
These findings are consistent with the view that the differential change in employment between treatment and control groups in the 1990s is explained by welfare reform, rather than

[3] Meyer and Rosenbaum [2001] analyze 1984–1996, Hoynes and Patel [2018] study 1991–1998, Strain and Schanzenbach [2021] examine 1989–1998, and Michelmore and Pilkauskas [2021] study 1990–2016. I replicate the results of these studies over a consistent 1991–2001 period except for Meyer and Rosenbaum [2001], where I analyze only 1991–1996.

[4] Summary statistics for this sample are provided in Appendix A.

[5] This rate is the same measure of predicted welfare exposure introduced above in connection with Figure 1.

Figure 2: Replication of Canonical EITC Specifications with and without Controlling for Time × Pre-Reform Welfare Use



Notes: "×" markers show baseline replication estimates; circles show estimates after adding Year × predicted pre-reform welfare use interactions. Vertical bars are 95% confidence intervals clustered at the state-by-year level. SWS = Strain and Schanzenbach [2021]; HP = Hoynes and Patel [2018]; MR = Meyer and Rosenbaum [2001]; MP = Michelmore and Pilkauskas [2021]. Sample subgroups: All = all single women; ≤HS = high school or less; <HS = less than high school; >HS = more than high school. MP specifications interact EITC with age of youngest child (0–2, 3–5, 6–12, 13+). Sample: Single women ages 20–50, March CPS 1991–2001.

by the EITC itself. Since WelfareExposure_i is a pre-treatment, time-invariant characteristic estimated only with 1991–1993 data and is not conditioned on post-reform behavior or contemporaneous policy take-up, interacting it with year dummies simply allows differential time paths proportional to baseline exposure, which is precisely the potential bias we seek to remove.

At the same time, one might worry that this approach effectively "over-controls" by soaking up part of the true effect of the EITC if the policy's impact happened to be correlated with baseline welfare exposure. Is Welfare Exposure a valid control that addresses confounding, or does it absorb changes that should be attributed to the EITC? The next section addresses this concern.

4 Differential Exposure to Welfare Reform as a Source of Bias

Including interactions between year and baseline welfare exposure eliminates the canonical EITC effects. A remaining concern, however, is that this specification introduces a "bad control," potentially absorbing part of the true EITC effect. To evaluate this concern, I show that the apparent labor-supply responses to the 1993 expansion are driven by differential exposure to welfare reform across groups, not by the EITC itself.

I demonstrate these patterns in two ways. First, I examine employment trajectories within subgroups defined by their predicted welfare exposure. Mothers with similar pre-reform exposure follow nearly identical paths regardless of EITC treatment status, implying that the differences across treatment and control groups arise entirely from composition. Second, I conduct placebo tests that assign artificial "treatment" and "control" labels to subgroups who in reality faced the same EITC schedule. These tests show that the canonical DiD estimator generates spurious effects that scale precisely with differences in pre-reform welfare exposure. Together, this evidence demonstrates that the conventional DiD estimates are not identified from genuine treatment–control contrasts, but instead reflect compositional imbalance in exposure to welfare reform. Appendix C provides a formal decomposition of the DiD estimator, showing explicitly how differences in subgroup composition mechanically generate the observed estimates.

4.1 Evidence of Non-Parallel Trends

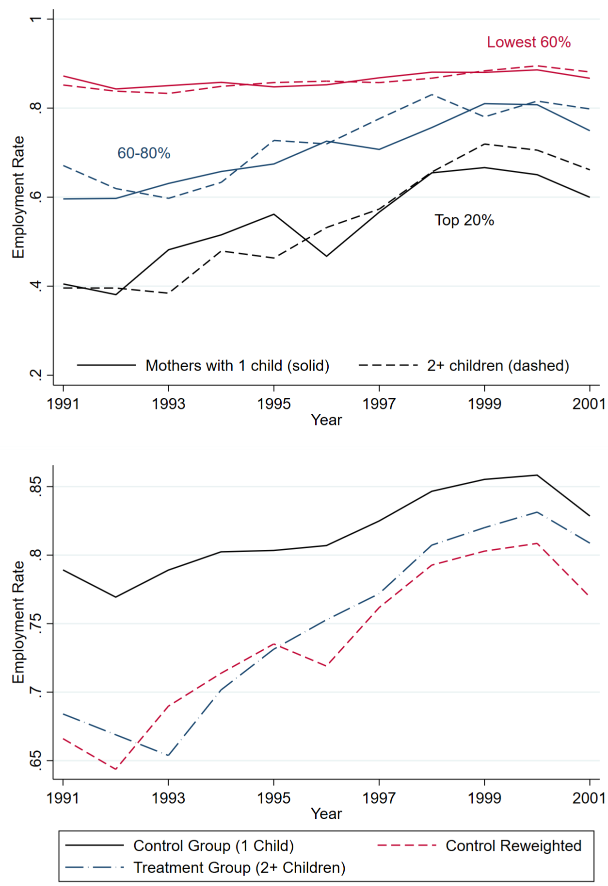
To assess whether the observed treatment effects reflect the EITC or differences in welfare exposure, I begin by comparing employment trends among women with similar predicted likelihoods of welfare use before reform. If the EITC were driving the results, then differences between treatment and control groups should persist even after conditioning on welfare exposure. By contrast, if exposure to welfare reform explains the results, then women with similar baseline propensities should follow parallel paths regardless of EITC eligibility.

The top panel of Figure 3 plots employment trends of single mothers, disaggregated by ex-ante predicted welfare participation (Lowest 60%, 60–80%, Top 20%), separately for mothers of one child (solid) and two or more children (dashed). The figure shows, first, that groups with higher baseline welfare participation experienced much larger employment increases after 1993, while groups with low baseline participation saw little change. Second, conditional on baseline exposure, mothers with one versus two or more children follow essentially the same trajectory. That is, the EITC treatment group and the control group are indistinguishable once matched on welfare exposure.

The second panel plots the employment rates of mothers of one child (the solid black line) and mothers of 2+ children (the blue long-dashed line) as well as mothers of one child reweighted based on their predicted pre-reform welfare exposure to match the welfare exposure of the treatment group. Rather than the large relative increase in employment of mothers of 2+ children, the lines follow the same path.

In other words, the composition of the two groups explains why conventional DiD estimates are large. Mothers with exactly one child are disproportionately drawn from the

Figure 3: Employment Rates of Single Mothers by Predicted Pre-Reform Welfare Use, 1991–2001



Notes: Top panel: Solid lines show mothers with one child; dashed lines show mothers with two or more children. Groups are stratified by quintiles of predicted pre-reform welfare exposure (bottom three quintiles combined as "Lowest 60%"). Conditional on baseline welfare exposure, employment trajectories are nearly identical across treatment and control groups. Bottom panel: The black line shows raw employment of mothers with one child (control); the red dashed line reweights this group to match the welfare-exposure distribution of mothers with two or more children; the blue dashed line shows raw employment of mothers with two or more children (treatment). After reweighting, the employment paths of treatment and control groups are nearly identical. Sample: Single mothers ages 20–50, March CPS 1991–2001.

lowest-exposure strata (77% in the lowest 60%, 16% in the 60–80%, and only 7% in the top 20%). By contrast, mothers with two or more children are much more concentrated in high-exposure strata (47% in the lowest 60%, 23% in the 60–80%, and 30% in the top 20%). Because employment surged only among high-exposure groups, and those groups are overrepresented in the treatment group, the difference-in-differences estimator mechanically attributes those gains to the EITC. Were the two groups balanced on pre-reform welfare exposure, the treatment effect of the EITC would essentially disappear, as Figure 3 shows.

4.2 Placebo Tests and Implementation

A complementary way to show that the canonical DiD estimator is biased is to run placebo tests that ask whether the estimator produces "effects" in contexts where, by construction, the true treatment effect is zero. If the estimator systematically generates spurious effects related to baseline welfare exposure, this shows that the identifying assumptions are violated and the failure to control for welfare exposure is the source of the bias. To see this formally, consider a simplified model

$$y_{it} = \beta \, \mathrm{EITC}_{it} + X_{it}\delta + \gamma Z_i + \varepsilon_{it},$$

where X_{it} are standard controls and Z_i is an omitted factor correlated with treatment. If Z_i is excluded, the estimated coefficient on EITC_{it} reflects omitted variables bias:

$$\hat{\beta} = \beta + \gamma \frac{\mathrm{cov}(\mathrm{EITC}_{it}, Z_i)}{\mathrm{var}(\mathrm{EITC}_{it})}.$$

When EITC_{it} is a simple post×treatment dummy, this expression reduces to

$$\hat{\beta} = \beta + \gamma \Delta Z, \tag{2}$$

that is, the effect of Z on y times the mean difference in Z across groups.

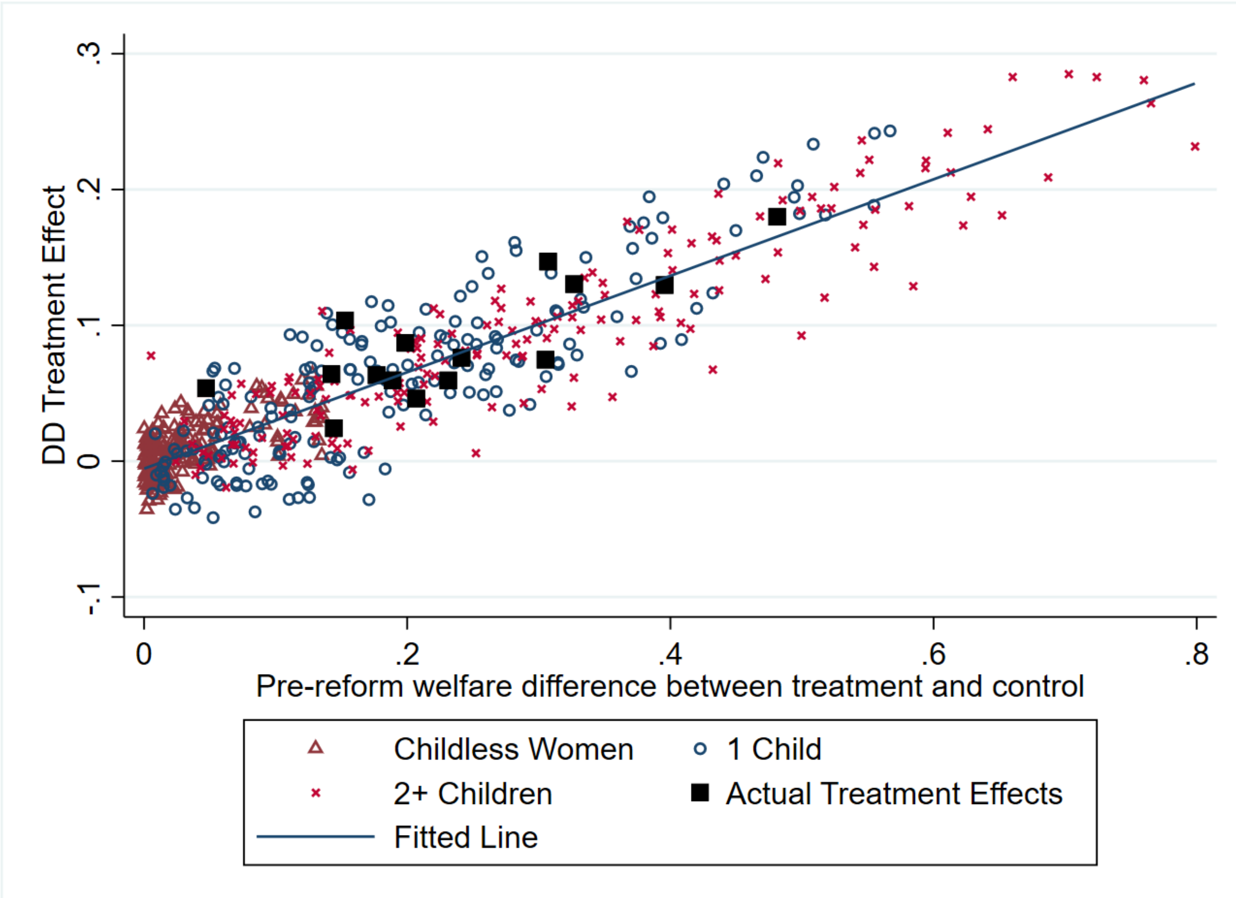
This framing clarifies the logic of the placebo design: if we construct groups where the treatment effect is the same (so that $\beta = 0$), any nonzero $\hat{\beta}$ must reflect bias. Moreover, if we find that placebo estimates are correlated with some (omitted) variable ΔZ, that reveals the source of the bias. Unlike the classic case of unobserved ability in wage regressions, here the relevant confounder—exposure to welfare reform—is measurable, but deliberately excluded under the assumption that other controls (e.g., state-by-year fixed effects) are sufficient.

Design and Implementation. I form treatment and control groups that are equally exposed to the EITC, so any estimated difference-in-differences must be spurious. To do this, I restrict attention to mothers with the same EITC treatment status (e.g., mothers with one child, mothers with two or more children, or childless women). Within each group, I divide individuals into subgroups based on pre-reform characteristics. In particular, using 1991–1993 data, I estimate predicted probabilities of (i) welfare receipt and (ii) nonemployment from probit models, and (iii) predicted log wages from OLS regressions, with covariates for education, race, marital status, age of youngest child, and state of residence. I then sort individuals into deciles of each predicted variable within each EITC treatment group.

For each EITC group, I assign one decile as a "placebo treatment" and another as a "placebo control," and re-estimate the DiD regression. Because these subgroups face the same EITC policy schedule, the true treatment effect should be zero. Repeating this procedure across all possible decile pairings yields 90 placebo estimates per outcome measure. For each regression, I also record the difference in predicted welfare exposure between the two placebo groups as a measure of imbalance in the suspected confounder.

I compare these placebo estimates to the actual DiD estimates from Strain and Schanzenbach [2021], Meyer and Rosenbaum [2001], Hoynes and Patel [2018], and Michelmore and Pilkauskas [2021]. Because each paper compares multiple groups (e.g., mothers of 2+ children, mothers of one child, and childless women), I form and estimate all the pairwise comparisons that comprise each estimate following Goodman-Bacon [2021]. These can then be directly compared to the placebo estimates.

Figure 4: Placebo and Actual DiD Estimates by Pre-Reform Welfare Exposure



Notes: Each hollow symbol represents a placebo DiD estimate comparing two decile groups within the same EITC treatment category: triangles for childless women, circles for mothers with one child, and "×" for mothers with two or more children. Deciles are formed by predicted pre-reform probability of welfare receipt and nonemployment. By construction, all placebo comparisons have a true treatment effect of zero. The horizontal axis shows the difference in pre-reform welfare exposure between the placebo "treatment" and "control" groups. Black squares are actual DiD estimates from pairwise comparisons in Strain and Schanzenbach [2021], Hoynes and Patel [2018], Meyer and Rosenbaum [2001], and Michelmore and Pilkauskas [2021]. The fitted line (slope = 0.41, s.e. = 0.02) is estimated on placebo estimates only. Sample: Single women ages 20–50, March CPS 1991–2001.

Results. Figure 4 shows two clear findings of this exercise. First, placebo estimates are often large and economically meaningful: the DiD estimator finds "effects" where none exist. Second, the magnitude of these spurious effects is linear in the pre-reform welfare-use gap between groups. When placebo groups are similar, estimates are near zero; when imbalanced, estimates are large. Importantly, the actual DiD estimates from the canonical literature fall exactly on this same fitted line.

In other words, knowing only the placebo results from any single treatment group and the baseline welfare-use gap is enough to predict the published results—without needing any data on post-reform outcomes from other treatment groups. Taken together, these findings imply that the canonical DiD estimates are not identified from genuine treatment–control contrasts in EITC exposure. They are mechanically generated by differences in baseline exposure to welfare reform.

4.3 Corroboration: Welfare Time Limits

A useful corroboration comes from a related literature on welfare time limits. Grogger [2004] estimates the effect of newly imposed five-year time limits on welfare use, exploiting variation in the age of the youngest child: families with younger children face a binding constraint (they could exhaust benefits before the child ages out), while families whose youngest child is a teenager do not. Crucially, this identification strategy relies on entirely different treatment variation than the EITC studies—age of youngest child rather than number of children—and the theoretical mechanism (precautionary savings of welfare benefits) is unrelated to EITC incentives. Yet the estimator is equally exposed to the same potential confound, because mothers of younger children also had higher baseline welfare use than mothers of teenagers.

Appendix F shows that the time-limit estimates exhibit the same pattern documented above. The DiD estimates fall on the same fitted line as both the EITC estimates and the placebo estimates (Figure 11), and controlling for Year $\times$ pre-reform welfare exposure attenuates them toward zero (Figure 12). The fact that an entirely separate identification strategy, based on different treatment groups and a different policy mechanism, produces the same spurious pattern reinforces the conclusion that the bias arises from differential exposure to welfare reform rather than from any particular policy's true effect.

5 Preventing Bias from Omitted Variables with Machine Learning

One way to prevent this type of error in future research is to systematically search for potential confounding factors ex ante, augmenting and complementing researchers' discretion in covariate selection.

A pragmatic approach is to use machine learning (ML) to search for omitted variables that may bias the estimator. Conventional approaches to covariate selection are often ad hoc, relying on researcher discretion about which variables to include or how to specify them. Even more systematic approaches, such as matching or synthetic control methods, typically rely only on pre-reform data and exclude information from the post-reform period. As a result, they may miss variables that play little role in pre-trends but become important confounders in the post-reform period (like exposure to welfare reform). A more systematic alternative is to let the data itself highlight which variables predict outcomes across groups.

I demonstrate this approach using the simple lasso algorithm. The lasso selects a sparse set of predictors from a large pool of covariates. In this case, I form a large set of potential baseline covariates interacted with year effects (such as demographics, predicted pre-reform welfare use, employment, and wages). Concretely, I train the lasso first on one group (say, mothers with one child) to predict employment outcomes over time, and then generate out-of-sample predictions for the other group (say, mothers with two or more children).

The crucial feature is that training within a single group withholds any information on treatment status. The cross-group predictions therefore represent the ML algorithm's best guess at what the untreated counterfactual path would have looked like.

The intuition is as follows: if the EITC truly caused the divergence between groups, then these cross-predictions should systematically miss, with predicted employment too high for controls and too low for the more-treated group. If instead the predictions track observed outcomes closely, that suggests the divergence is fully explained by observable confounders—such as baseline welfare exposure—that the lasso has identified as predictive.

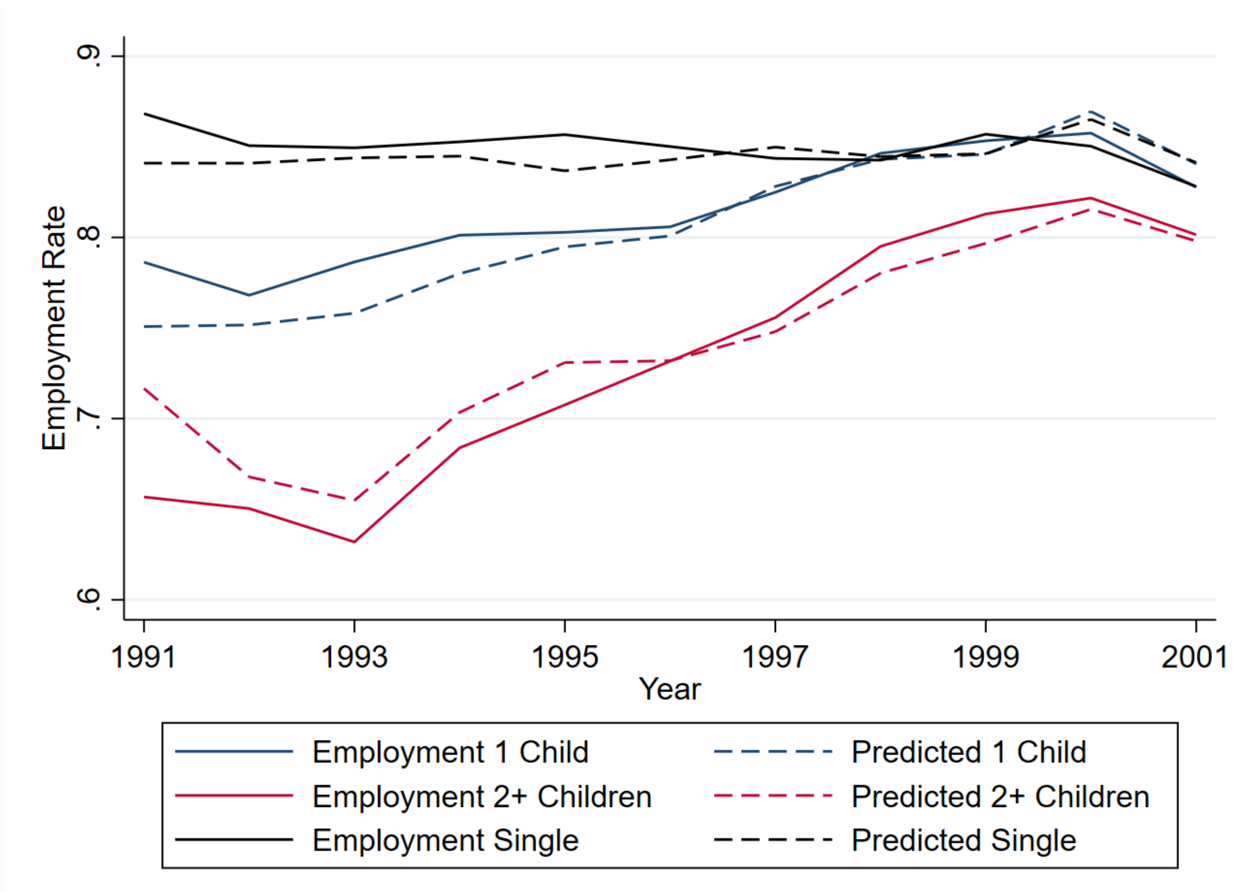
This approach therefore lends itself directly to the familiar graphical assessment of the parallel trends assumption: by plotting actual and predicted outcomes side by side, one can visually assess whether treatment and control groups would have followed similar paths absent treatment. Figure 5 illustrates this, comparing actual employment trends (solid lines) with the out-of-sample ML predictions (dashed lines). Because predictions for each group are generated exclusively from data on a different group, they contain no information on treatment status and thus serve as an untreated counterfactual. In this case, the exercise illustrates that the identifying assumptions are unlikely to hold.

First, Figure 5 shows that the out-of-sample predictions closely track the actual outcomes. This is evidence against a large EITC treatment effect: the counterfactual constructed from the other group matches realized outcomes remarkably well. If the EITC were the primary driver, we would expect large, systematic forecast errors—predicted employment that is too high for controls and too low for the more-treated group. Instead, the alignment between actual and predicted outcomes reinforces the conclusions from earlier sections and provides a direct graphical check of the parallel trends assumption.

Second, the ML procedure consistently selects interactions of time with pre-reform welfare participation (or correlates of welfare participation, like age of the youngest child) important predictors within groups. These are the same confounders tied to exposure to welfare reform identified in Sections 3 and 4. With these covariates in hand, one could have generated the correct specification ex ante or, at minimum, flagged that the parallel-trends assumption was unlikely to hold post-reform.

Finally, the cross-predictions serve as a transparent design diagnostic. Even when pre-trends appear parallel, the post-reform out-of-sample predictions make clear that treatment and control groups were unlikely to remain parallel absent treatment—both because higher-welfare-use subgroups exhibit distinct within-group trends and because employment among childless women was already near its ceiling.

Figure 5: Actual and Predicted Out-of-Sample ML Employment Rates



Notes: Solid lines show actual annual employment rates; dashed lines show out-of-sample predictions from a lasso model with cross-validation. Covariates include demographics, age polynomials, and interactions of year with predicted pre-reform welfare use, employment, and wages. Predictions for mothers of two or more children and childless women are estimated exclusively from data on mothers of one child; predictions for mothers of one child are estimated from data on mothers of two or more children. Because predictions are trained on a different group, they contain no information about treatment status and represent the algorithm's best guess of counterfactual outcomes. The close alignment between actual and predicted outcomes leaves little room for an EITC treatment effect. Sample: Single women ages 20–50, March CPS 1991–2001.

6 Conclusion

This paper demonstrates that the canonical difference-in-differences designs used to evaluate the 1993 EITC expansion are biased by differential exposure to welfare reform—an observable, pre-determined characteristic that differs systematically between treatment and control groups. Three pieces of evidence support this conclusion. First, replication with a single control—$\text{Year}_t \times \text{WelfareExposure}_i$—collapses the canonical EITC estimates to zero. Second, reweighting and placebo tests reveal that the estimated effects arise entirely from compositional imbalance rather than from genuine treatment–control contrasts in EITC exposure. Third, a simple ML cross-prediction flags the violation ex ante and surfaces welfare exposure as the omitted covariate. Applied to the EITC literature, these results imply that welfare reform—not the EITC expansion—drove the large employment gains of single mothers in the 1990s.

Ethnographic accounts emphasize shifts in the culture and administration of welfare—pressures to reduce caseloads and tighten eligibility. The empirical results here show that exposure to these reforms, proxied by predicted pre-reform welfare use, explains most of the change in employment over the decade. Because the 1993 expansion underpins much of the EITC literature, the implications extend beyond labor supply. Studies of poverty [Hoynes and Patel, 2018], child achievement [Dahl and Lochner, 2012], maternal health [Evans and Garthwaite, 2014], infant health [Hoynes et al., 2015], and other outcomes may likewise misattribute to the EITC effects actually caused by welfare reform.

These findings also bear on current policy debates by underscoring that employment responses in the 1990s were associated with reductions in unconditional support, not solely with work-contingent subsidies.

References

Rebecca M. Blank. Evaluating welfare reform in the united states. *Journal of Economic Literature*, 40(4):1105–1166, December 2002. doi: 10.1257/002205102762203576. URL https://www.aeaweb.org/articles?id=10.1257/002205102762203576.

Gordon B. Dahl and Lance Lochner. The impact of family income on child achievement: Evidence from the earned income tax credit. *American Economic Review*, 102(5):1927–56, August 2012. doi: 10.1257/aer.102.5.1927. URL https://www.aeaweb.org/articles?id=10.1257/aer.102.5.1927.

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

William N. Evans and Craig L. Garthwaite. Giving mom a break: The impact of higher eitc payments on maternal health. *American Economic Journal: Economic Policy*, 6(2): 258–90, May 2014. doi: 10.1257/pol.6.2.258. URL https://www.aeaweb.org/articles?id=10.1257/pol.6.2.258.

Sarah Flood, Miriam King, Renae Rodgers, Steven Ruggles, J. Robert Warren, Daniel Backman, Annie Chen, Grace Cooper, Stephanie Richards, Megan Schouweiler, and Michael Westberry. IPUMS CPS: Version 12.0 [dataset], 2024. URL https://doi.org/10.18128/D030.V12.0.

Andrew Goodman-Bacon. Difference-in-differences with variation in treatment timing. *Journal of Econometrics*, 225(2):254–277, 2021. doi: 10.1016/j.jeconom.2021.03.014.

Jeffrey Grogger. Time limits and welfare use. *The Journal of Human Resources*, 39(2): 405–424, 2004. ISSN 0022166X. URL http://www.jstor.org/stable/3559020.

V. Joseph Hotz and John Karl Scholz. The earned income tax credit. Working Paper

8078, National Bureau of Economic Research, January 2001. URL http://www.nber.org/papers/w8078.

Hilary Hoynes, Doug Miller, and David Simon. Income, the earned income tax credit, and infant health. *American Economic Journal: Economic Policy*, 7(1):172–211, February 2015. doi: 10.1257/pol.20120179. URL https://www.aeaweb.org/articles?id=10.1257/pol.20120179.

Hilary W. Hoynes and Ankur J. Patel. 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, 2018.

Hilary W. Hoynes and Diane Whitmore Schanzenbach. Safety net investments in children. *Brookings Papers on Economic Activity*, Spring:89–132, 2018.

Henrik Kleven. The eitc and the extensive margin: A reappraisal. *Journal of Public Economics*, 236:105135, 2024. ISSN 0047-2727. doi: https://doi.org/10.1016/j.jpubeco.2024.105135. URL https://www.sciencedirect.com/science/article/pii/S0047272724000719.

Lawrence M. Mead. Overselling the earned income tax credit. *National Affairs*, 21, 2014. Fall.

Bruce Meyer and Dan Rosenbaum. Welfare, the earned income tax credit, and the labor supply of single mothers. *Quarterly Journal of Economics*, 116(3):1063–1114, 2001.

Katherine Michelmore and Natasha Pilkauskas. 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, 2021.

Austin Nichols and Jesse Rothstein. *The Earned Income Tax Credit*, pages 137–218. University of Chicago Press, November 2016. URL http://www.nber.org/chapters/c13484.

Michael R. Strain and Diane Whitmore Schanzenbach. Employment effects of the earned income tax credit: Taking the long view. *Tax Policy and the Economy*, 35(1):87–129, 2021.

R. Kent Weaver. *Ending Welfare as We Know It*. Brookings Institution Press, 2000. ISBN 9780815792475. URL http://www.jstor.org/stable/10.7864/j.ctvdmwzqn.1.

A Summary Statistics

The analysis uses data from the March Current Population Survey (CPS) Annual Social and Economic Supplement, obtained via IPUMS [Flood et al., 2024], for the years 1991–2001. The sample is restricted to unmarried women ages 20–50 (excluding childless women enrolled in school). Mothers whose youngest child is older than 18 are dropped, as are observations with implausible age gaps between the mother and her youngest child. Income variables are deflated to constant dollars using the CPI.

The primary outcome is an indicator for any positive earnings in the previous year. Covariates include a quartic polynomial in age, indicators for education (less than high school, high school, some college, college or more), race, marital status, age of youngest child, and state of residence. All regression specifications include state-by-year fixed effects.

EITC treatment variables are constructed following each of the four replicated studies and merged from external sources: the Strain and Schanzenbach [2021] post-1993 treatment indicator (with phase-in adjustments), the Hoynes and Patel [2018] simulated EITC instrument (incorporating state EITC programs), the Meyer and Rosenbaum [2001] average taxes-if-work variable by number of children, and the Michelmore and Pilkauskas [2021] average EITC by state, year, and number of children. State welfare time-limit indicators are merged separately for the analysis in Appendix F.

Table 1 reports summary statistics for the pooled sample. Table 2 shows annual means of key outcomes and treatment variables by number of children.

Table 1: Summary Statistics

Variable	Mean	SD
Earned any income	0.82	0.38
Welfare participation	0.10	0.29
Earnings ($)	20,898	25,630
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	150,432	

Table 2: Panel Outcomes and Policy Variables by Year

	1991	1992	1993	1994	1995	1996	1997	1998	1999	2000
Panel A: Childless Women										
Work	0.87	0.85	0.85	0.85	0.86	0.85	0.84	0.84	0.86	0.85
Welfare	0.03	0.04	0.03	0.02	0.02	0.02	0.02	0.01	0.01	0.01
SWS Treatment	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				
HP Sim EITC	0	0	0	40	40	40	40	41	40	39
Panel B: Mothers with One Child										
Work	0.79	0.77	0.79	0.80	0.80	0.81	0.82	0.85	0.85	0.86
Welfare	0.23	0.24	0.24	0.22	0.19	0.18	0.15	0.12	0.10	0.08
SWS Treatment	0	0	0	0.92	1	1	1	1	1	1
MR Tax if Work	1083	1001	955	702	633	609				
HP Sim EITC	625	693	729	957	1010	1013	924	948	941	917
MP Average EITC	711	781	825	1092	1155	1160	1172	1204	1193	1166
Panel C: Mothers with Two or More Children										
Work	0.66	0.65	0.63	0.68	0.71	0.73	0.76	0.80	0.81	0.82
Welfare	0.38	0.37	0.39	0.35	0.31	0.29	0.25	0.20	0.16	0.14
SWS Treatment	0	0	0	0.50	0.78	1	1	1	1	1
MR Tax if Work	651	554	501	19	−291	−548				
HP Sim EITC	633	707	745	1194	1457	1670	1440	1473	1469	1436
MP Average EITC	780	865	916	1453	1769	2014	2028	2079	2060	2024

B Replication and Comparison to the Literature

B.1 Data and Sample

The analysis uses the March CPS from 1991–2001, restricting to non-married women ages 20–50 who are not enrolled in school. Mothers whose youngest child is older than 18 are excluded. The dependent variable is an indicator for employment last year (positive earnings). Covariates include a quartic in age, dummies for education, race, age of youngest child, and marital status. All specifications include state-by-year fixed effects.

B.2 Treatment Variables

I focus on four influential papers that are representative of the published evidence on the EITC's labor supply effects, use common and well-known data from the March CPS, but differ in their parameterization of the EITC, the subpopulations studied, and whether they incorporate only federal or both federal and state EITC expansions.

Strain and Schanzenbach [2021] (SWS). The EITC expansion is represented as a dummy variable equal to zero before 1994; equal to 0.92 for mothers of one child and 0.5 for mothers of two or more children in 1994 (reflecting partial phase-in); equal to 1 for mothers of one child and 0.78 for mothers of two or more children in 1995; and equal to 1 for both groups thereafter. Childless women serve as the control group. Policy variation is at the national level, and the authors report estimates for subpopulations by educational attainment (less than high school, high school only).

Meyer and Rosenbaum [2001] (MR). The EITC's effect is captured through the variable "taxes if work," which measures the average level of taxes (net of the EITC) if working, varying by year and number of children (0, 1, or 2+). This specification exploits variation at the year $\times$ number-of-children level and was estimated over the period 1984–1996, specifically excluding post-1996 welfare policy changes.

Hoynes and Patel [2018] (HP). The EITC is represented as a simulated dollar value in each state (incorporating any state-level EITC) by year and by number of children (0 to 3). This approach captures both federal expansions and cross-state variation from state EITC programs. The authors report estimates for the full sample and by educational attainment.

Michelmore and Pilkauskas [2021] (MP). This study examines heterogeneity in the EITC's effect by the age of the mother's youngest child. The EITC is modeled as a simulated value at the state $\times$ year $\times$ number-of-children level (for mothers with 1–3 children; childless women are excluded), and this treatment variable is interacted with dummies for age of youngest child (0–2, 3–5, 6–12, 13+). The analysis spans 1990–2016.

B.3 Baseline Replication

Figure 6 compares my re-estimated EITC effects to the published coefficients across these four studies. Red dots are my replication; blue crosses are the published literature; vertical bars are 95% CIs for the replication. Despite differences in sample periods, treatment parameterization, and subsamples, the replication closely matches the published magnitudes and statistical significance patterns.

With the exception of Meyer and Rosenbaum [2001], each specification is estimated over the period from 1991-2002. The replication of Meyer and Rosenbaum [2001]'s specification is 1991-1996.

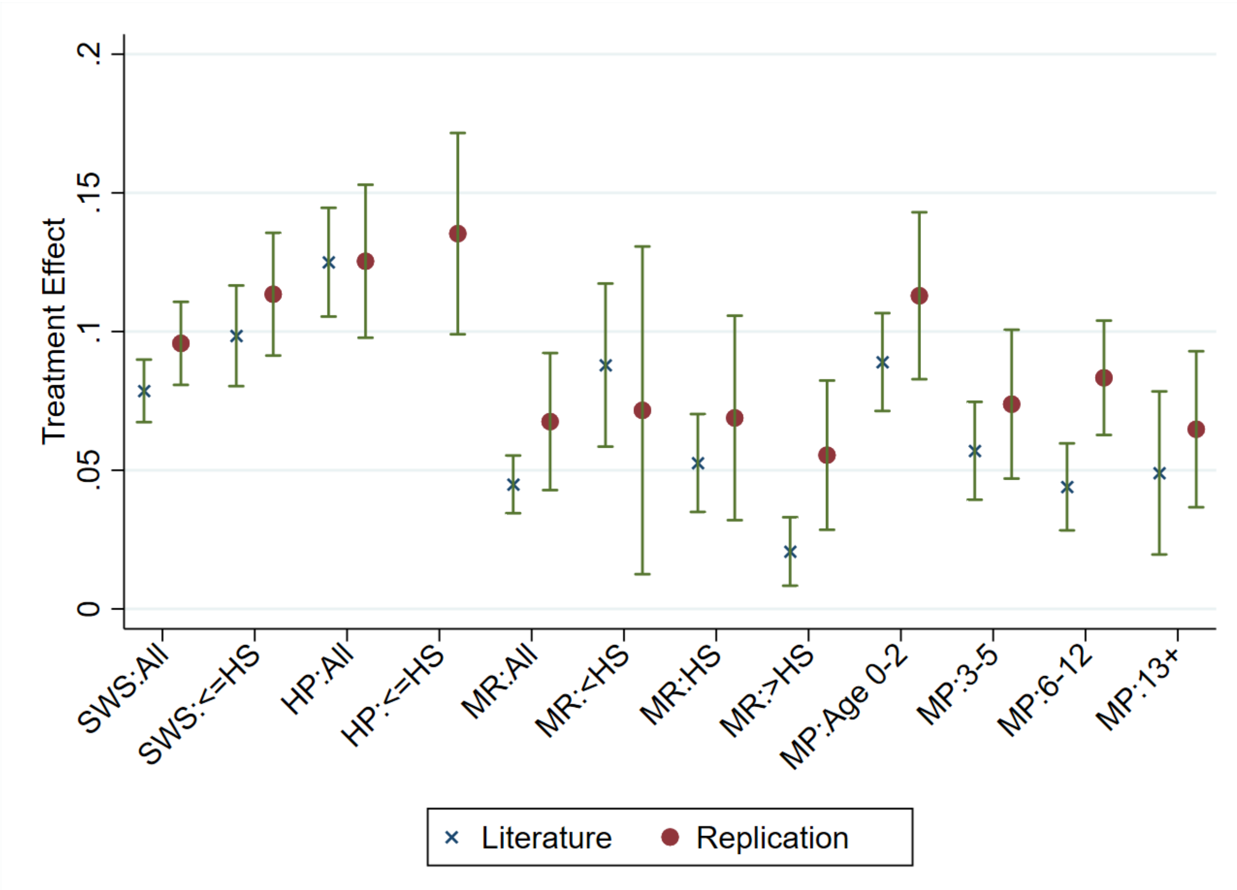
Whereas Michelmore and Pilkauskas [2021] use "worked last week" as the outcome, my replication uses "earned any income last year"—yet the scales remain comparable. Overall, the replication corroborates the canonical estimates and thus provides a credible baseline for the controlled specifications that add $\text{Year}_t \times \text{WelfareExposure}_i$. Table 3 reports the underlying coefficient estimates and standard errors.

Table 3: Replication of Published EITC Estimates on Employment

	Literature	Replication
Panel A: Strain and Whitmore Schanzenbach		
All women	0.0786	0.0957
	(0.0058)	(0.0076)
$\leq$ High school	0.0985	0.1134
	(0.0093)	(0.0113)
Panel B: Hoynes & Patel		
All women	0.1250	0.1253
	(0.0100)	(0.0141)
$\leq$ High school		0.1353
		(0.0185)
Panel C: Meyer & Rosenbaum		
All women	0.0449	0.0675
	(0.0053)	(0.0126)
$<$ High school	0.0879	0.0716
	(0.0150)	(0.0301)
High school	0.0526	0.0688
	(0.0090)	(0.0188)
$>$ High school	0.0207	0.0554
	(0.0063)	(0.0137)
Panel D: Michelmore & Pilkauskas		
Age 0–2	0.0890	0.1129
	(0.0090)	(0.0154)
Age 3–5	0.0570	0.0738
	(0.0090)	(0.0137)
Age 6–12	0.0440	0.0833
	(0.0080)	(0.0105)
Age 13+	0.0490	0.0648
	(0.0150)	(0.0143)

Notes: Each row reports a point estimate and standard error (in parentheses) from the published literature and my replication. Panel A: post-1993 EITC treatment dummy (2+ children $\times$ post). Panel B: simulated EITC instrument ($1,000s). Panel C: taxes-if-work variable ($1,000s); sample restricted to years before 1997. Panel D: EITC maximum benefit interacted with age-of-youngest-child group dummies. All regressions include demographic controls (age quartic, education, race, marital status, number of children, age of youngest child), state fixed effects, and year fixed effects, weighted using CPS survey weights with standard errors clustered at the state $\times$ year level.

Figure 6: Published vs. Replicated EITC Treatment Effects



Notes: Blue "×" markers are point estimates reported in the published literature; red dots are my replication with 95% confidence intervals (vertical bars). SWS = Strain and Schanzenbach [2021]; HP = Hoynes and Patel [2018]; MR = Meyer and Rosenbaum [2001]; MP = Michelmore and Pilkauskas [2021]. Specifications differ in EITC parameterization (dummy, simulated value, taxes-if-work), sample restrictions (by education), and whether state EITC programs are incorporated. All replications use March CPS 1991–2001, except MR which uses 1992–1997. Standard errors clustered at the state-by-year level.

B.4 Controlling for Welfare Exposure

Table 4 reports the results of adding Year × predicted pre-reform welfare exposure interactions to each specification. Across all studies and subsamples, the EITC coefficients attenuate toward zero, while the welfare × year interaction is large, positive, and highly significant—indicating that the original estimates were driven by differential exposure to welfare reform rather than the EITC.

Table 4: EITC Estimates Controlling for Pre-Reform Welfare Exposure

	(1) EITC Coefficient	(2) Welfare × Year Coefficient	(3) Mean Welfare Exposure	(4) Implied Δ Employment	(5) Actual Δ Employment
Panel A: Strain and Whitmore Schanzenbach					
All women	0.0114	0.5650	0.095	0.054	0.055
	(0.0071)	(0.0510)			
≤ High school	0.0107	0.5637	0.166	0.094	0.094
	(0.0123)	(0.0606)			
Panel B: Hoynes & Patel					
All women	0.0062	0.5789	0.095	0.055	0.055
	(0.0122)	(0.0497)			
≤ High school	-0.0248	0.6090	0.166	0.101	0.094
	(0.0188)	(0.0613)			
Panel C: Meyer & Rosenbaum					
All women	0.0082	0.2719	0.095	0.026	0.024
	(0.0139)	(0.0509)			
< High school	-0.0028	0.2742	0.285	0.078	0.025
	(0.0426)	(0.1151)			
High school	-0.0484	0.6054	0.114	0.069	0.032
	(0.0263)	(0.1236)			
> High school	0.0231	0.4128	0.039	0.016	0.009
	(0.0168)	(0.1080)			
Panel D: Michelmore & Pilkauskas					
Age 0–2	0.0052	0.6116	0.351	0.215	0.230
	(0.0170)	(0.0543)			
Age 3–5	0.0050		0.208	0.127	0.124
	(0.0147)				
Age 6–12	0.0164		0.167	0.102	0.112
	(0.0119)				
Age 13+	0.0087		0.093	0.057	0.057
	(0.0145)				

Notes: Columns (1)–(2) report coefficient estimates and standard errors (in parentheses) from regressions that add Year × predicted pre-reform welfare exposure interactions to the baseline specifications in Table 3. Column (1): EITC coefficient after controlling for welfare exposure. Column (2): coefficient on the interaction of predicted pre-reform welfare use with the post-period year indicator (1999 for Panels A, B, and D; 1996 for Panel C); regressions include the full set of year interactions but only the key post-period interaction is shown. Column (3): weighted mean of predicted welfare exposure (Pr(on welfare and not working | X)) in each specification's sample. Column (4) = (2) × (3): the implied average employment increase attributable to welfare reform for the sample. Column (5): actual change in the employment rate from 1993 to the post period (1999 or 1996). See Table 3 notes for other specification details.

C Decomposition of the DiD Estimator

This appendix formalizes why compositional differences between treatment and control groups can bias the DiD estimator even when subgroup-specific trends are identical across groups.

Consider the simplest 2×2 DiD setup, where superscripts t and c index treatment and control groups and subscripts 0 and 1 index time before and after treatment. The DiD estimator is:

$$\hat{\beta} = (\bar{y}_1^t - \bar{y}_0^t) - (\bar{y}_1^c - \bar{y}_0^c) = \Delta\bar{y}^t - \Delta\bar{y}^c.$$

Now suppose individuals belong to subgroups s with potentially different time trends. Let ω_s^i denote the share of group i in subgroup s, and let Δy_s^i denote the change in outcomes for subgroup s within group i. The group-level changes can be written as weighted averages:

$$\Delta\bar{y}^t = \sum_s \omega_s^t \Delta y_s^t, \qquad \Delta\bar{y}^c = \sum_s \omega_s^c \Delta y_s^c.$$

Adding and subtracting $\sum_s \omega_s^t \Delta y_s^c$ (the control group's subgroup trends weighted by treatment group composition) yields:

$$\hat{\beta} = \underbrace{\sum_s \omega_s^t \left(\Delta y_s^t - \Delta y_s^c\right)}_{\text{Within-subgroup treatment effect}} + \underbrace{\sum_s \left(\omega_s^t - \omega_s^c\right) \Delta y_s^c}_{\text{Compositional bias}}.$$

The first term is the weighted average of subgroup-specific treatment effects. The second term is the compositional bias: it captures how differences in the distribution of subgroups across treatment and control groups, combined with differential subgroup trends, generate spurious effects.

Under the parallel trends assumption, the second term should be zero. This requires either (1) subgroup trends are identical (Δy_s^c is constant across s), or (2) the composition of treatment and control groups is balanced ($\omega_s^t = \omega_s^c$ for all s).

In the EITC context, neither condition holds. Mothers with two or more children are disproportionately drawn from high-welfare-exposure subgroups, and those subgroups experienced much larger employment increases. As a result, the compositional bias term is large and positive, fully accounting for the observed DiD estimate.

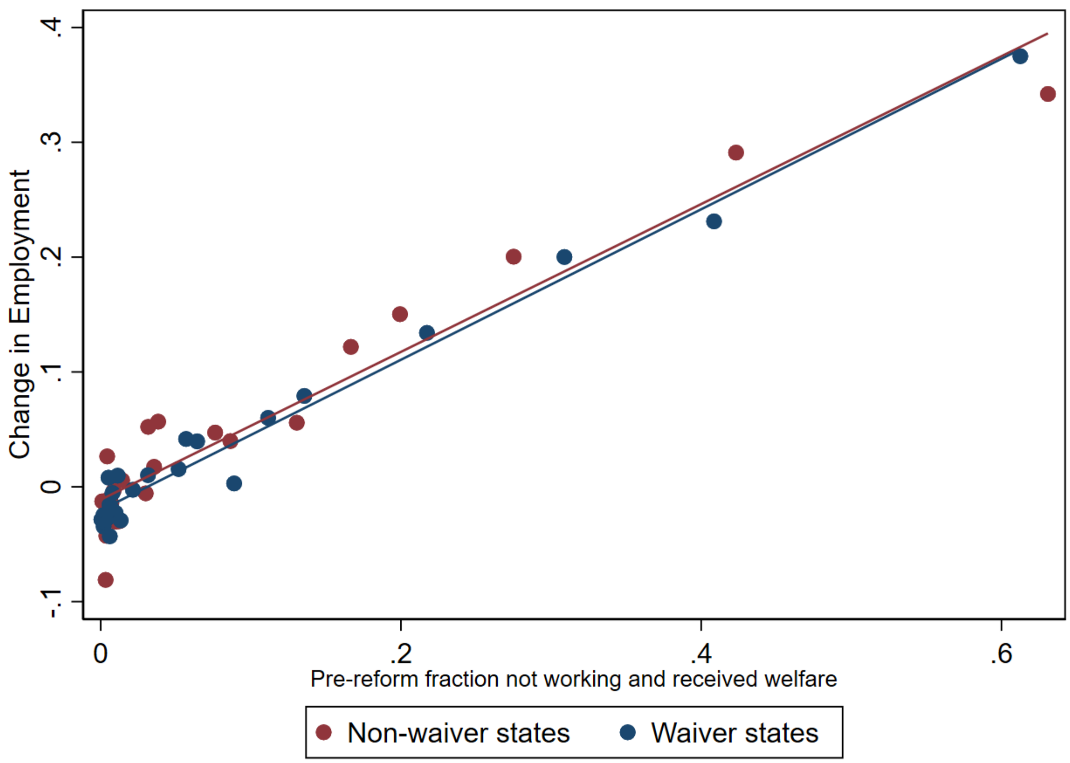
D Waiver vs. Non-Waiver States

An important robustness check is whether the proportional-exposure pattern varies across waiver and non-waiver states. Strain and Schanzenbach [2021] stress that AFDC waivers generated policy heterogeneity in the early 1990s and argue that including state×year fixed effects accounts for confounding from welfare reform. If the bias I document were solely a product of waiver states, one would expect systematically different relationships in waiver versus non-waiver states.

Figure 7 plots the change in employment from 1993 to 2000 against the pre-reform share of women not working and receiving welfare, separately for waiver and non-waiver states. The slopes are essentially identical, indicating that the proportional-exposure pattern is national, not driven by waiver timing.

This evidence reinforces two points. First, the proportional-exposure mechanism is not confined to waiver states but holds nationally. Second, controlling for state×year effects, as in Strain and Schanzenbach [2021], does not eliminate the bias because the imbalance in exposure occurs within states as well as across them.

Figure 7: Employment Changes by Pre-Reform Welfare Exposure: Waiver vs. Non-Waiver States



Notes: Each point represents a bin of single women grouped by predicted pre-reform welfare exposure, plotted separately for states that implemented AFDC waivers before 1996 (navy) and states that did not (maroon). The horizontal axis shows pre-reform welfare exposure; the vertical axis shows the change in employment between 1993 and 2000. The fitted lines are nearly identical (slopes of 0.62 and 0.67, respectively), indicating that the proportional-exposure pattern is national and not driven by waiver timing. Sample: Single women ages 20–50, March CPS 1991–1993 and 1999–2001.

E 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 differences 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.

E.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.

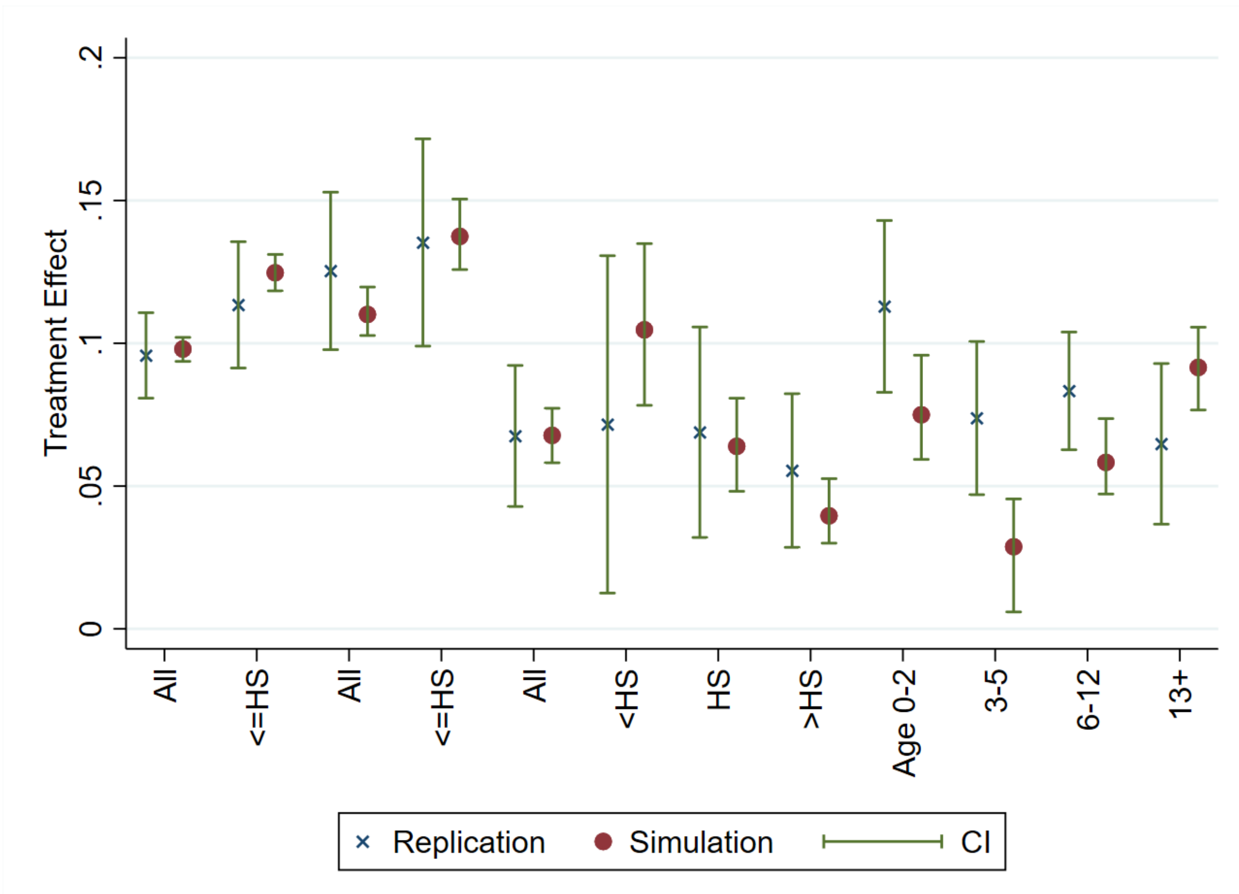
I estimate the same regression specifications as above. The results, presented in Figure 8, yield two findings.

First, the TWFE coefficient estimates from the simulated data closely match those from the actual analysis, despite the fact that simulated welfare exits and job entries are assigned entirely at random with no differential treatment by number of children, education, or any other characteristic. The implication is that an equal proportional reduction in welfare use—consistent with the administrative changes and work requirements described in the ethnographic literature—mechanically generates the same DiD estimates as those reported in the canonical studies. Put differently, the regression specifications cannot distinguish between the EITC and a simple, uniform caseload reduction.

Second, the simulation rationalizes several "puzzles" in the data that have been cited as evidence for or against the EITC interpretation. Kleven [2024], for example, highlights the fanning-out of employment gains by family size and the disproportionately large responses among groups with the highest baseline welfare use as anomalies that are difficult to reconcile with EITC incentives alone. Figures 9 and 10 show that the random-exit simulation accurately predicts these patterns. Figure 9 plots welfare receipt and employment by number of children (0 through 4+), comparing actual trends (solid lines) to simulated trends (dashed lines). The simulation reproduces both the sharp decline in welfare use and the corresponding increase in employment across all groups, including the pronounced fanning-out among mothers with three or four or more children. Figure 10 presents the same comparison by age of youngest child, using the categories from Michelmore and Pilkauskas [2021] (0–2, 3–5, 6–12, 13+). Here too the simulated trends closely track the actual data: the largest declines in welfare and the largest increases in employment occur among mothers with the youngest children—precisely the groups with the highest baseline welfare exposure—and the simulation matches these differential changes without any heterogeneous treatment.

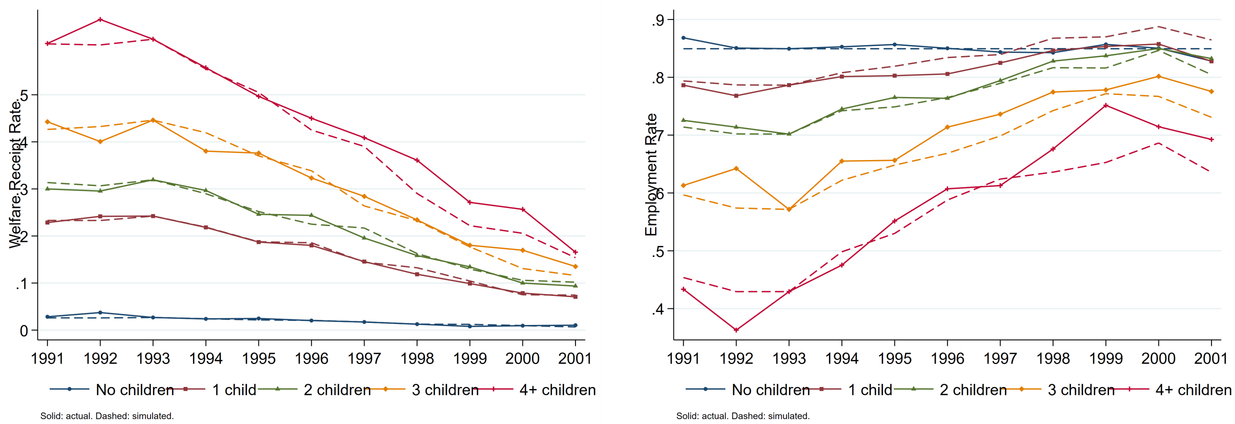
The close correspondence between simulated and actual outcomes across both dimensions corroborates the conclusion that the employment gains of the 1990s are fully consistent with a uniform, proportional reduction in welfare caseloads. Because the simulation contains no EITC variation whatsoever, its ability to replicate both the regression estimates and the subgroup-level trends provides strong evidence that welfare reform, not the EITC, drove the observed changes.

Figure 8: Simulation vs. Replication: Estimated Effect of EITC



Notes: "×" markers show EITC coefficient estimates from baseline replications in the March CPS. Circles show estimates from simulated data in which the only time-series variation comes from randomly removing mothers from welfare rolls and assigning them to employment at rates matching aggregate CPS trends—with no differential treatment by number of children or other characteristics. The simulation reproduces the canonical DiD estimates, demonstrating that an equal proportional reduction in welfare use (consistent with the caseload reduction hypothesis) mechanically generates the observed treatment effects. Vertical bars show the range of estimates across 500 Monte Carlo draws. Sample: Single women ages 20–50, March CPS 1991–2001.

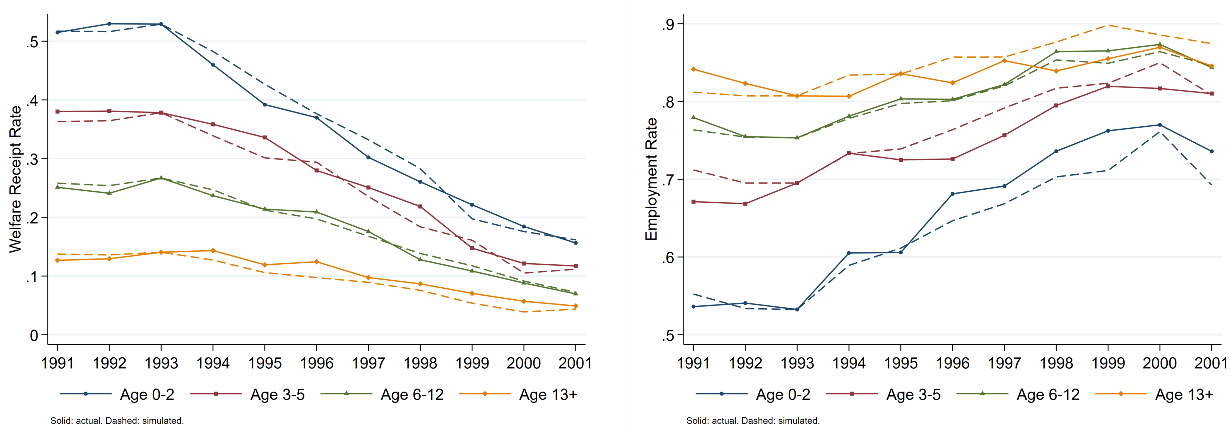
Figure 9: Actual vs. Simulated Welfare Receipt and Employment by Number of Children



Welfare Receipt　　　　Employment

Notes: Solid lines with markers show actual annual means from the March CPS; dashed lines show means from one draw of the simulation in which welfare exits and job entries are assigned uniformly at random to match aggregate CPS trends. No differential treatment is applied by number of children. The simulation reproduces the fanning-out pattern across family sizes. Sample: Single women ages 20–50, March CPS 1991–2001.

Figure 10: Actual vs. Simulated Welfare Receipt and Employment by Age of Youngest Child



Welfare Receipt　　　　Employment

Notes: Solid lines with markers show actual annual means from the March CPS; dashed lines show simulated means. Age-of-youngest-child categories follow Michelmore and Pilkauskas [2021]: 0–2, 3–5, 6–12, and 13+. Mothers only (restricted to women with at least one child). The simulation reproduces the heterogeneous changes by age of youngest child without any differential treatment. Sample: Single mothers ages 20–50, March CPS 1991–2001.

F Time Limits

In a related literature, 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 confounding effect on estimates of welfare time limits. Moreover, finding a confounding 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 "0–6" and "7 to 12") plus dummy variables for the age of youngest child.

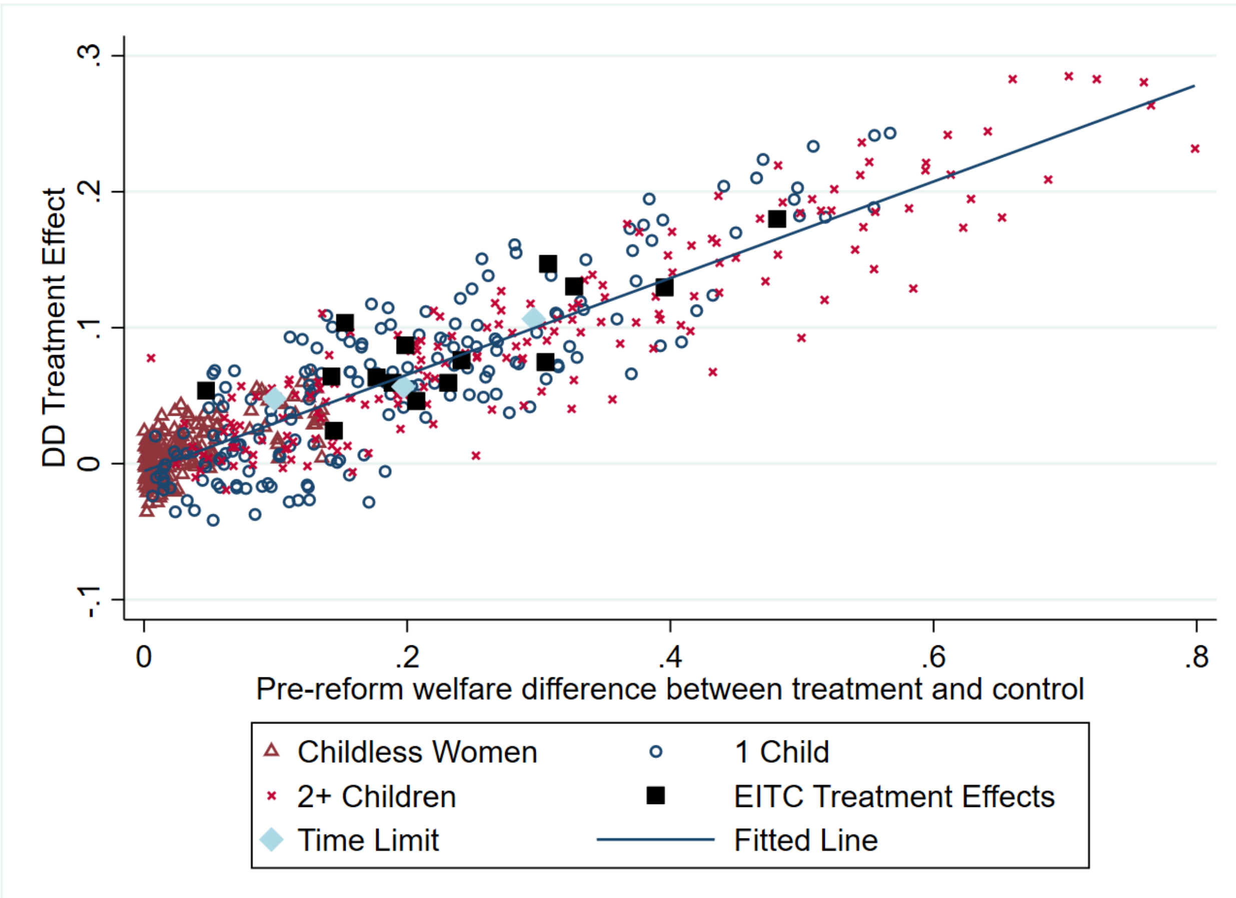
F.1 Analysis

In my reanalysis of time limits, I replicate the same general approach as I take to the EITC. First I reproduce 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 uses the simplest specification from Grogger (2004) for clarity, though the results 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.

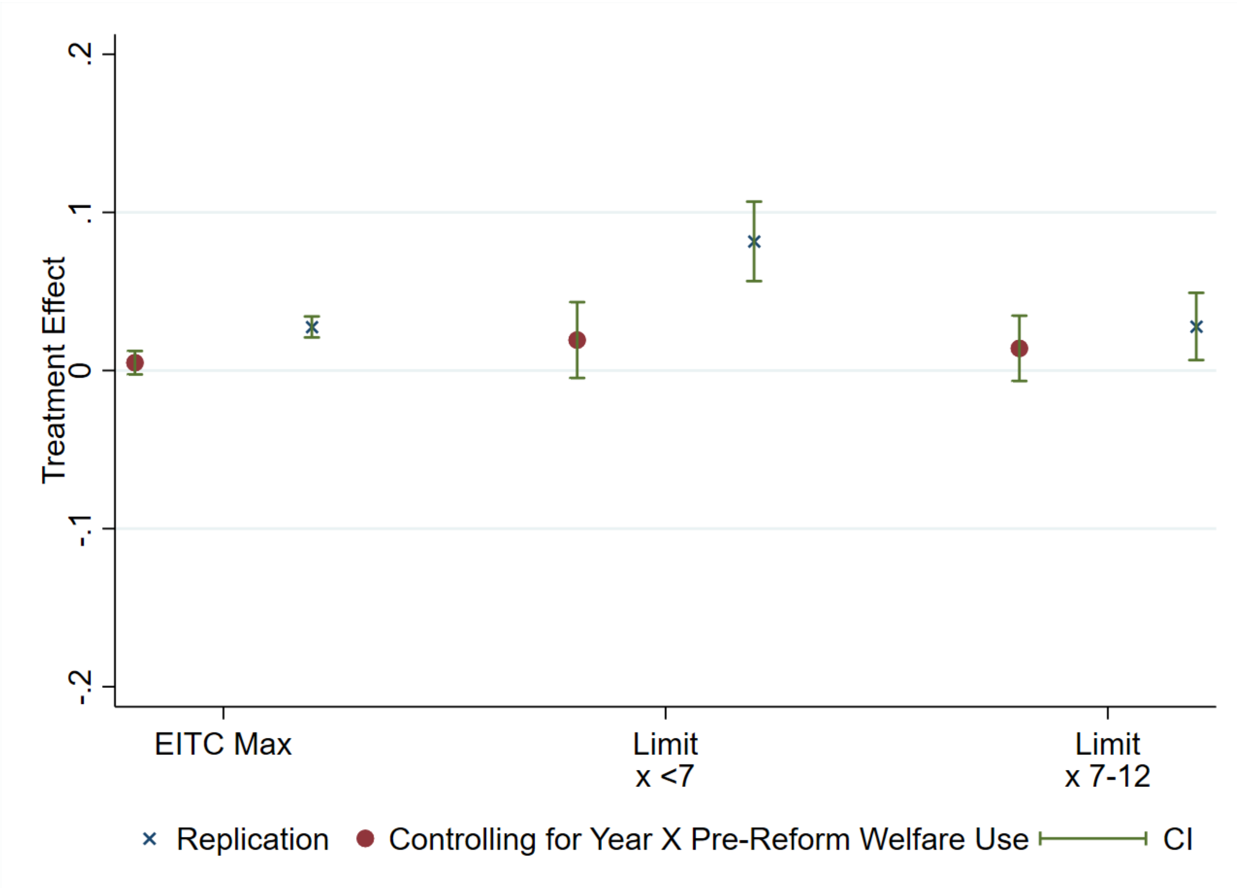
Finally, I include controls for ex-ante welfare use by time effects—the same Year $\times$ predicted pre-reform welfare exposure interactions used in Section 3—and show that the estimated effect of time limits on employment is attenuated toward zero, just as the EITC effects were.

Figure 11: Placebo and Actual DiD Estimates: EITC and Welfare Time Limits



Notes: Hollow symbols show placebo DiD estimates as in Figure 4. Black squares are actual EITC treatment effect estimates from pairwise comparisons in the published literature. Blue diamonds are DiD estimates of the effect of welfare time limits following Grogger [2004], comparing mothers whose youngest child is under 13 (affected by time limits) to mothers whose youngest child is 13–17 (unaffected because welfare eligibility ends at 18). Time limit estimates fall on the same fitted line as EITC estimates, indicating that both reflect the same underlying confound: differential exposure to welfare reform. Sample: Single women ages 20–50, March CPS 1991–2001.

Figure 12: Time Limit Estimates With and Without Welfare Exposure Controls



Notes: Estimates of the effects of the EITC ("EITC Max") and welfare time limits ("Limit × <7" and "Limit × 7–12") on employment, following Grogger [2004]. "×" markers show baseline replication estimates; circles show estimates after adding Year × predicted pre-reform welfare use interactions. Vertical bars are 95% confidence intervals. As with the EITC specifications, controlling for welfare exposure attenuates the time limit coefficients toward zero. Sample: Single mothers ages 20–50 with youngest child under 18, March CPS 1991–2001.