

Cutoff From Support: The Effects of Losing Cash Welfare*

Most Recent Version

Motaz Al-Chanati Lucas Husted[†]
Columbia University Columbia University

November 1, 2021

Abstract

Does removing families from welfare programs result in increased employment? Using detailed administrative data from Michigan, we study a policy reform in the state's TANF program that swiftly and unexpectedly removed over 10,000 families from welfare while quasi-randomly assigning time limits to over 30,000 remaining participants. Consistent with economic theory, removing families from welfare increases formal labor force participation by roughly 4 percentage points (20% over control group mean), with increases in annualized earnings of roughly \$500. We find that time limits – particularly for those further from their expiration – and sanctions due to program noncompliance yield similar results. However, despite this, the majority of families remain formally unemployed after welfare removal, and using quantile regressions we show that even the highest percentile wage gains fail to offset the loss in welfare benefits. Overall, our findings provide evidence that, contrary to their stated goals, welfare reform measures that either kick families off welfare or make welfare harder to access are likely to deepen poverty.

JEL Codes: H53, H75, I30, I38, J38

***Acknowledgments:** We thank Sandra Black, Michael Best, Miguel Urquiola, Douglas Almond, Wojciech Kopczuk, Bentley MacLeod, Suresh Naidu, and all of the members of the Columbia applied microeconomics colloquium for their feedback and support. Special thanks go to Katie Zeiter (MDHHS) and Jasmina Camo-Biogradlija (MERI) for their support in data acquisition, alongside all of the staff at MDHHS, MDE, and MERI who made this research possible. This material is based upon work supported by the National Science Foundation under Grant No. 2018387. Additional support came from the Program for Economic Research (PER). Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation or Columbia University.

[†]Contact Information: lfh2119@columbia.edu

1 Introduction

The efficacy of social safety net programs in alleviating poverty, and their potential work (dis)incentives, has long been the subject of public debate in the United States, and an extensive academic literature reflects this. While this debate peaked in the mid-1990s when reformers – concerned that programs discouraged work – redesigned cash welfare to encourage labor force participation, in the last decade many states have aimed to impose tighter SNAP and Medicaid work requirements using similar policy arguments.¹ More generally, discussions of the labor market implications of the safety net – spurred by unemployment insurance expansions and cash bequests during the COVID-19 pandemic – will continue to influence public policy for decades to come.

Economic theory generally predicts that welfare programs result in a lower labor supply, but the empirical evidence is mixed, in part due to the structural barriers low-income individuals face in finding stable employment. While some evidence suggests that the overhaul of the existing cash welfare programs in 1996 led to an increase in labor force participation among recipients, it is difficult to separately identify the simultaneous changes that a strong economy, EITC, and welfare reforms had on the labor force participation of single mothers over the last 30 years ([Meyer and Rosenbaum, 2001](#); [Chan, 2013](#); [Eissa and Nichols, 2005](#); [Bitler et al., 2006](#); [Kleven, 2019](#)).

As a result, we still don't fully understand whether there are labor supply effects of losing cash assistance or to what extent frictions prevent individuals from finding employment without benefits. What happens when you remove families from welfare unexpectedly? How do time limits or work requirement affect participants' overall usage and benefit spells? Finally, what happens to the children of families on welfare? Do reforms that reduce or eliminate the generosity of benefits come at the cost of their human capital development due to reduced parenting capacity or increased financial precarity?²

Using detailed administrative data, this paper answers all of these questions by exploiting an unexpected change to the Temporary Assistance for Needy Families (TANF) program in Michigan. In the fall of 2011, the Michigan government enacted a law to tighten and retroactively enforce time limits on its TANF program ([French, 2012](#)). From one month to the next over 10,000 families in excess of the legislated time limits immedi-

¹See [Blank \(2002\)](#), [Moffitt \(2003\)](#), or [Ziliak \(2015\)](#) for comprehensive reviews of AFDC/TANF. For books on the welfare reform see [Grogger and Karoly \(2006\)](#) or [Ziliak \(2009\)](#). For recent coverage of policy changes see [Galewitz \(2018\)](#), [Garfield et al. \(2019\)](#), or [Reiley \(2019\)](#).

²Future versions of this work will include analysis of the educational attainment of the children whose parents are removed from welfare (data linkage approved and forthcoming). This draft currently only addresses the immediate impacts of the policy reform studied on the labor supply outcomes, and ignores the children.

ately lost their benefits. Those who remained on the program were assigned time limits in accordance with their prior usage and – over the next five years – 30,000 families had their cases closed as they reached these limits (Lawler, 2016). This policy change thus provides quasi-random variation needed to explore the effect that welfare reforms have on the lives of current and former participants.

Our contribution is twofold. First, we add credible causal estimates of the labor supply effects of losing welfare to the existing literature. While we are not the first to address this topic, our context and data allow us to draw a more complete picture of the financial state of families who have lost benefits. The reform we study is more recent and free from confounding policy reforms, like changes to the EITC, that are endemic to past work. In particular, prior research has been limited by the fact that major welfare changes were implemented at approximately the same time across all states, making it difficult to trace out the effects that specific reforms had on the finances of this population (Eissa and Hoynes, 2004; Grogger, 2003; Grogger et al., 2003; Chan, 2013).³ This study is the only one to our knowledge that involves cases where some families faced an immediate negative wealth shock from a welfare reform – as opposed to diminished benefits or harsher program rules alone – and others in the same population did not.

Second, since the policy creates quasi-random variation in the time limit assigned to participants, we can directly estimate the effects of such limits on welfare use and employment. There has been considerable work examining the effects of time limits on program participation (Chan, 2013; Grogger, 2003; Grogger and Michalopoulos, 2003; Grogger, 2004; Meyer and Rosenbaum, 2001; Swann, 2005). In contrast to this existing literature, the variation we observe is free from confounding changes and does not require calibration of a structural model to estimate the effects of the limits. We observe individuals throughout our sample with similar characteristics but heterogeneous time limit assignments. While prior approaches provide rich insights about welfare use, they are still limited by the simultaneity of reforms during the time studied (Kleven, 2019). The variation in time limit length assigned here could not have been forecasted prior to the policy and varies across program participants.⁴

³EITC expansions and the addition of time limits and work requirements on welfare changed many things at once about the returns to working. Simultaneously, the US economy was in an economic expansion at the time that these changes were made, naturally removing families from welfare rolls.

⁴The variation in Grogger and Michalopoulos (2003) is due to the age of youngest child based on entry into the welfare program. So if you begin using welfare when your child is 15, you can only use 3 years of benefits before your eligibility horizon ends. Conversely, if your youngest child is 3, then you must allocate 5 years of benefits over 15 potential years of eligibility. This variation at the time of the program implementation of time limits creates variation to identify the effects of time limits. By contrast, our study semi-randomly (conditional on past use) allocates usable welfare stock to people regardless of the age of

Our findings are nuanced. Beginning with the labor market consequences of losing TANF unexpectedly, we find that single women with dependent children increase their (formal) labor market participation by 4 percentage points in response to welfare removal, an increase of nearly 20% over the control group mean of 25%. Removal increases the average (annualized) earnings of these families by roughly \$500, representing a 25% increase over the control group mean. These results are relatively large – given existing literature and labor force participation of poor families – and statistically robust to a variety of counterfactual empirical exercises.

However, these gains come at a high cost to the overall indigent population on TANF. Our findings indicate that the majority (60%) of these families remain unemployed with no formal labor market earnings after losing benefits. Using unconditional quantile regressions, we find that conditional on working, the 95th percentile of gains in formal wages fall below the previous benefit allotment. These findings combined show that families who are kicked off of TANF are, on average, poorer as a result of losing benefits. Thus removing women with children from welfare – or placing onerous restrictions on their welfare use – does not succeed in making them financially independent, but more likely makes them disconnected from both the welfare state and the formal labor market (Blank and Kovak, 2009; Turner et al., 2006).

We use the quasi-random assignment of time limits to the remaining TANF population to quantify the effect that such limits have on welfare use and employment. Consistent with the findings of Grogger and Michalopoulos (2003) – who show that those women further away from the end of their eligibility horizon face more incentive to bank benefits as insurance for negative wage draws in the future – we find that time limits cause about 10% of families to leave welfare, even far from their expiration. This treatment effect grows monotonically the closer families are to the complete exhaustion of benefits.

Time limits may affect families differently depending on the length of time over which they plan to use welfare, since the same number of months of available benefits stretched over a longer time horizon means that families should be more cautious with their use (Grogger and Michalopoulos, 2003). We test this logic formally by calculating the elasticity of welfare use and employment with respect to an additional month of welfare eligibility. We find that a 1% percent decrease in the welfare eligibility window leads to a 0.18% decrease in welfare use over the entire post period. This subsequently leads to a 0.15% increase in employment and a 0.31% increase in wages over the three years following the policy. Effects are largest for younger women with the longest benefit horizons.

We next explore the effects of welfare sanctions – policies that remove families from their children.

welfare temporarily for failing to meet program requirements – on program participation and labor supply.⁵ While the self-selection of individuals into noncompliance limits our ability to make strictly causal claims about sanctions, we find compelling evidence that such policies have large participation and employment effects. In particular, following a short spell (3 months) of loss of welfare eligibility, 60% of sanctioned families are still off welfare one year later. These findings imply that even minor administrative burdens can lead to permanent impacts on program participation, a likely explanation for the rise in disconnected mothers (Blank, 2007; Blank and Kovak, 2009; Turner et al., 2006). As with those individuals removed due to the reform, many sanctioned heads-of-household find employment after losing benefits. At the highest point, they make \$500-\$800 more per quarter compared to before the sanction. However, just as with those ultimately removed from welfare permanently by the policy change, these women are not doing better off financially compared to their prior benefit levels.

As a final suggestive exercise, we estimate event studies comparing those who timed out “naturally” – those who were aware that their benefits were soon-to-expire – to those who timed out due to the unexpected policy change; the latter are in a more precarious position, with earnings significantly lower than those removed with warning three years later. This implies that warning individuals ahead of time that their benefits will expire (thus allowing them time to adjust) can blunt the impacts of reforms to programs that provide cash assistance to the poor.

This work ties directly to the large literature that studies welfare programs, specifically those that involve direct cash assistance or the loss of welfare benefits. Several contributions study state-level TANF natural-experiments around the time that the AFDC transitioned to TANF (see Ziliak (2015) for a review). Notably Grogger and Michalopoulos (2003) study the effect of time-limits on welfare use in Florida and Kline and Tartari (2016) study a welfare reform experiment in Connecticut. Our study resembles these seminal works in that it will use a state-specific policy change to answer fundamental questions about the economic incentives of welfare.

While there is an extensive literature studying the labor supply effects of the AFDC as well as the experimental programs piloted under state waivers, the literature on TANF and cash welfare is far more limited. Notable exceptions include Bennett et al. (2004), Low et al. (2018), and Bitler et al. (2006). Our project relates directly to recent structural work in understanding the dynamic nature of the choices recipients face (Chan, 2013; Swann, 2005; Keane and Wolpin, 2010) and other work about similar economic incentives with

⁵In some ways this tests the effect of a particular feature of work requirements on welfare use and employment.

different welfare programs (Garthwaite et al., 2014; Hoynes and Schanzenbach, 2012). We build upon work by Deshpande on the Supplemental Security Income (SSI) program, by studying the long-term financial consequences of removing individuals from welfare (Deshpande, 2016a,b). This paper further speaks to the growing literature that is starting to revisit some of the programs that were significantly changed in the late 1990s, or are currently being overhauled, with new data or empirical methods (Kleven, 2019; Gray et al., 2021).

The rest of this paper is organized as follows. Section 2 provides a background on the Michigan TANF program including the reform we study. Section 3 details our data (3.1) with summary statistics and our empirical strategy (3.2). Section 4 provides the main empirical results in detail. Section 5 details robustness of the main empirical results with further discussion in the appendix. Section 6 concludes. Finally, this draft is preliminary in nature. Aside from key results that may differ from later versions, the primary difference between this draft and later ones is that it does not contain any analysis of the children affected by the reform policy. Subsequent versions will contain more data and outcomes of interest specifically related to these family members indirectly affected by the policy change.

2 Background and Policy Change

2.1 Temporary Assistance for Needy Families (TANF)

The TANF program – also known as “cash welfare” – is the only liquid cash benefit given to non-disabled families in the United States. TANF is the successor of the Aid to Families with Dependent Children (AFDC) program, arising out of the Personal Responsibility and Work Opportunity Act (PRWORA) of 1996, known colloquially as “welfare reform.” The new welfare program was no longer an entitlement, and the funding mechanism, fixed block-grants, gave states broad flexibility in program administration. While TANF continued to give cash payments to families, it included several additional goals related to family formation and work that were meant to end “welfare dependence.” The most important and controversial of these changes were adding time limits, work requirements, and sanctions to welfare to move participants into the labor force.⁶

The total TANF budget is roughly \$32 billion in combined federal (~\$16.5 billion) and

⁶The overall goals of the TANF legislation were to provide assistance to needy families, end welfare dependence by promoting work, prevent out of wedlock pregnancies, and encourage the formation of two parent families.

state spending.⁷ Due to the fact that the grants and earnings disregards are not indexed to inflation, the economy has generally expanded, and work requirements and sanctions have moved people off the program, the number of families receiving cash transfers has declined substantially since the 1990s. States have also repurposed block grants away from direct aid. With approximately the same nominal funding, in 1996 70% of funds went directly to welfare payments, while today the number is around 20%. States vary wildly in the amounts they provide.⁸ Many states spend less than 10% on direct aid, and overall, nearly half of all funds are used to supplement state Earned Income Tax Credits (EITCs) or pre-Kindergarten education (Schott et al., 2015).⁹

However, for those receiving cash benefits, it is a vital lifeline. The average 3-person family (1 adult, 2 children) – representative of the typical TANF recipient in Michigan – receives about \$500 per month (Michigan DHHS, 2011). Moreover, less than 20% of Michigan TANF recipients in 2011 had non-TANF earnings (U.S. DHHS, 2013). This means many families rely entirely on TANF and SNAP, and are still living below the federal poverty line. It is worth noting that a typical 3-person family with no earnings would receive just over \$500 in SNAP benefits; therefore, losing cash assistance can be viewed as equivalent to losing 50% of income for these families (Michigan DHHS, 2011). As noted by Ziliak (2015), the overall TANF population is younger and less likely to be in the labor force now as compared to the population in the 1990s, suggesting that the current population is more vulnerable than in the past. Despite its importance for these families, TANF has received less attention by academics and the public as compared to other social safety programs.¹⁰

2.2 Michigan's TANF Program

As of 2012, to qualify for Michigan's TANF program a family must have liquid assets lower than \$3000 and monthly income minus disregards less than the maximum benefit

⁷States are required to chip in with state funds based on spending for the AFDC program and related programs prior to TANF.

⁸For example, in 2014 California spent 46% of its funding on direct aid with a maximum of benefit of \$670 for a family of three, while Texas spent 7% with a maximum benefit of \$277.

⁹States often use the additional goals of TANF related to family formation and the encouragement of work to divert funds away from direct cash payments to needy families.

¹⁰Ziliak (2015) provides a comprehensive recent review of the literature. As noted there and elsewhere, there is scant evidence about the effects of welfare on the well-being of participants.

(based on family size).¹¹¹² Specifically, the monthly benefit is calculated as:

$$\text{Benefit} = \max \{ G_k - \mathbb{1} \{ Y \geq 200 \} \times 0.5 \times (Y - 200) - N, 0 \} \quad (1)$$

where G_k is the maximum monthly benefit for a family of size k , Y is monthly earned income, and N is any other countable income, like court-ordered child support from a family member not currently in the home.¹³ The income of children under age 18 is excluded entirely. The earnings disregard for the benefit reduction is \$200; after this benefits face a tax of 50% (the central expression).

In order to receive benefits, eligible recipients must work. If their youngest child is under (over) six, single parents are required to work 20 (30) hours a week, while two parent households are required to work 35 (55) combined hours a week. Various activities – besides normal unsubsidized earnings – count towards this requirement, including subsidized employment, on-the-job training, job search, unpaid work, and community service. Individuals are not required to work if they are over 65, a disqualified alien, a caretaker for a child who continues to receive benefits (a type of child-only case), or caring for a spouse with disabilities.

Upon entering TANF, participants are referred to service centers where they work with caseworkers to connect them with employment opportunities.¹⁴ These caseworkers not only connect them with jobs, but also refer them to additional support that can lower the barriers to employment, such as child care or transportation services. Michigan operates several programs to lower the cost of hiring TANF participants, including work subsidy schemes with participating employers.

If an eligible participant fails to meet work requirements or else does not attend other mandatory sessions, they are given a sanction which results in non-payment of welfare for a span of several months. Termination of benefits can occur after multiple sanctions

¹¹Michigan's TANF-funded cash welfare program is called the Family Independence Program (FIP). To avoid confusion when discussing federal programs, we refer to the FIP as Michigan's TANF program.

¹²This excludes vehicles and real estate. In 2012, families could have a home valued at up to \$200,000. These amounts have changed slightly over time.

¹³For an eligible grantee, G_k starts at \$306 for a single person, and goes up by around \$100 a person for each family member added. It is \$492 for a family of three and \$597 for a family of five. Additionally, the test for benefit eligibility is based on an 80% earnings disregard rate, while the benefit is based on a 50% earnings disregard rate. This change was made in 2011, alongside the time limit rules, to make the benefit slightly more generous. In practice, this changed nothing for people not in the labor force, while giving employed families an extra (small) nominal amount each month.

¹⁴There has always been such a program in Michigan over the period studied. Prior to 2013, this program was called JET (Jobs, Education and Training), and it is now called PATH (Partnership. Accountability. Training. Hope.).

have occurred.¹⁵ Recipients can receive (temporary) deferral from work if they are under age 16, have a child younger than 2 months, have a short term disability (or are being evaluated for one) that limits their ability to participate in work, meet SSI disability criteria, have a doctor verified excuse, or are a victim of domestic violence.

Participants receive welfare funds through state or federal sources. The source of funding can matter for state administrators, since it allows them to set distinct priorities for cash eligible groups separate from federal guidelines, but is immaterial to recipients.¹⁶ Each month that participants are on TANF counts towards their state and federal time limits unless they are ineligible due to a permanent or temporary deferral.¹⁷ Notably, sanctioned months count towards the limit even if the family does not receive funding.¹⁸

From 2002 through 2012, Michigan's TANF program had a stable caseload of approximately 70,000 to 80,000 cases per month. The average grant amount and case size was \$418 and 2.7 people, respectively. The modal case is one in which there is one mother and two children; indeed, 92% of cases had a woman as the grantee and only ~7% of cases had two parents. Most families do not spend long on the program. In the summer of 2011, the average number of months on assistance among all participants was 15 months, with only 16% of the families exceeding 24 months of use. The average age of the grantee was 30 years old and 55% of the grantees were black. Though federal rules require that a minimum of 50% of participants have to be involved in work-related activities to meet work requirements, in 2011 the rate in Michigan was closer to 20%, excluding excepted groups. By 2014, this number had reached 60%, indicative of changes after the policy encouraging more work participation, though this includes new training programs and other transitional programming ([Michigan DHHS, 2014](#)).

2.3 The 2011 Michigan Reform

While time limits for welfare benefits have been a feature of the TANF program since its inception in 1996, states can supplement the program with their own funding through segregated or separate funds or partially exempt recipients from work requirements or

¹⁵The first penalty is 3 months, the second 6 months, while the final is complete removal. Penalties are assessed on the individual but apply to the group, so it is possible for one family members' noncompliance to trigger removal for the family unit. We discuss this in more detail in section 4.3.

¹⁶This feature explains the key policy reform of this study. Michigan essentially converted federally ineligible recipients to state TANF funds. Michigan also sets aside state funding for married couples requiring benefits, among other priority groups.

¹⁷Two time limits are recorded since the funding source for the recipient changes depending on the type of deferral or exception being granted.

¹⁸Child-only cases and cases with other ineligible grantees as the primary recipients do not have time limits.

time limits altogether ([Moffitt, 2003](#); [Ziliak, 2015](#)). As of 2011, Michigan and 3 other states offered unlimited benefits out of state funds (with some restrictions), while 5 additional states time-limited the adult and provided funding for the child only when the limits were reached ([Ziliak, 2015](#)).¹⁹ Federal criteria also allows states to exempt up to 20% of TANF-funded cases from time limits due to hardship ([Hetling et al., 2020](#); [Ziliak, 2015](#)).²⁰

After a Republican state government came into power, the Michigan legislature passed a bill in September 2011 (effective October 1, 2011) to retroactively enforce previously enforced time limits. Since the TANF program had existed for over a decade and many families had already reached the federal 60 month limits while continuing to receive TANF without disruption, the policy meant that over 10,000 single-parent families currently enrolled would suddenly lose access to benefits. The act was touted as a way for the state to save money; estimates at the time projected 10,897 immediate case closures with an approximately 3,000 additional closures by the end of 2012 due to families reaching the limits. The savings were projected at roughly \$75 million ([Carley, 2011](#)).²¹ Cases that were child-only or where the recipient had a disability that incapacitates them were exempted from the policy.

Using our administrative data, figure 1 shows the effect of the policy on average caseloads in our sample. The first panel presents overall cases broken down by family structure. In October of 2011, the single parent TANF cases in the state of Michigan drop from over 50,000 to below 40,000. Two parent cases are relatively rare and seem unaffected, likely because fewer of these reach the 60 month limit to begin with. By design, child-only cases – not subject to time limits and work requirements – are unaffected. These cases comprise those in which children are eligible, but the parent or legal guardian – if present – is ineligible to receive assistance due to Supplemental Security Income (SSI) receipt, immigration status, or other program rules like sanctions.²² By extending to 2020, this figure also illustrates the general decline in cases and the shift from a population of single-parent to child-only cases on TANF.

The second subfigure plots the three most common reasons that individuals are re-

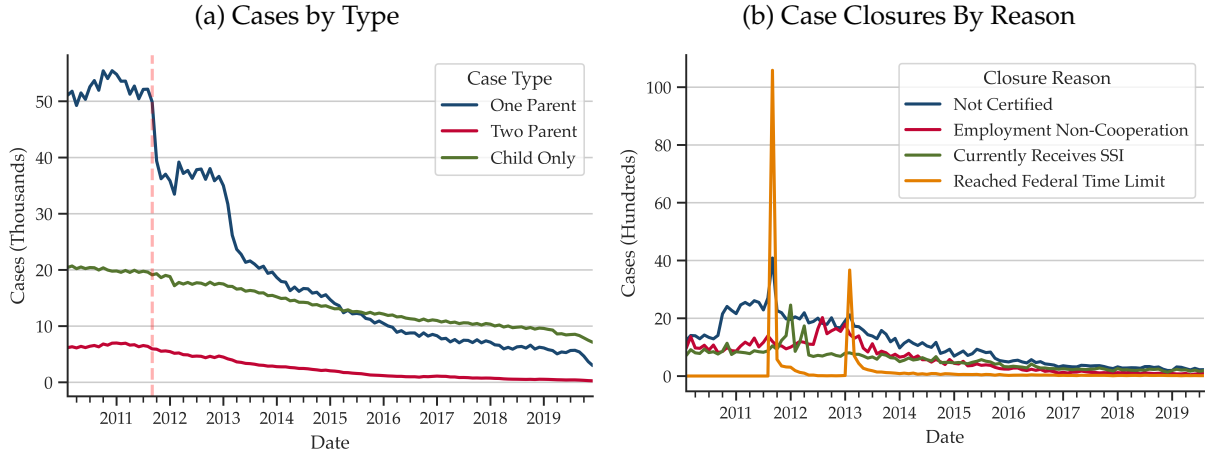
¹⁹As of 2012, Massachusetts, New York, Vermont and the District of Columbia all continued to pay benefits with no limit. California, Indiana, Maryland, Oregon, and Texas continued to provide benefits for the child-only case. There is other heterogeneity with respect to time limits. See [Ziliak \(2015\)](#) for more information.

²⁰Unsurprisingly, there is heterogeneity in the use of such exemptions with some states accounting for the vast majority of exemptions granted.

²¹The bill passed in the wake of the financial crises when many states were still attempting to close major budget shortfalls by re-evaluating their use of TANF funds ([Lower-Basch, 2011](#)).

²²See [Golden and Hawkins \(2011\)](#) for a high level review including links to studies of child-only cases.

Figure 1: Dynamics of TANF Cases in Michigan



Notes: All depicted lines show monthly summations. The first figure shows the number of cases per month in our sample broken down by case-type. Child only cases are ones in which the child is the eligible grantee and their guardian is a legal guardian (who may or may not be their biological parent). Two parent cases are subject to slightly different rules and mostly paid for with state-funded benefits. The second panel shows the three most common case closure reasons in our sample of case closures alongside the counts of those reaching the federal time limit. There were no instance of reaching the federal time limit prior to October 2011.

moved from TANF alongside timeout due to reaching the 60 month limit. The three most common reasons are a catchall term noting that the person is no longer certified; non-cooperation with employment rules resulting in (often) temporary sanction; and moving from TANF to SSI receipt, both of which cannot be claimed simultaneously. Removal due to timeout is fairly rare, comprising only the tenth most common reason for removal over the time examined. In our dataset that goes back to 2008, we have no evidence that any families were removed due to reaching the time limits prior to the reform in question. While it is possible that some of these families were included in the non-certified category – illustrated by the fact that the non-certified category also rises in September of 2011 – the evidence leads us to confirm that time limits were unenforced prior to 2011.

Of note is that the number of TANF cases plateaus and then plummets again in February of 2013. Due to the controversial nature of the reform, a lawsuit put a halt to certain removals, and even allowed some families to reapply for benefits in June of 2012. This stay on the new policy was officially removed at the start of 2013, and these families were formally removed then. This lawsuit thus attenuates the effects of the policy, though even families that reapplied for benefits and were approved would have spent nine months without them and at most have had seven months back on the program before being removed again. It may thus be more appropriate to think about the reform as having two waves.

In total, on October 2011, according to our data, there were 10,289 fewer families re-

ceiving benefits than in the month before (though more became permanently ineligible), or 38,553 fewer people when totaling the primary recipients and their children. In March, 2013, 5,775 additional cases were closed. The population removed was the poorest subsection of the TANF population. They also tended to have more kids on average. From September to October 2011, the number of children per case on one-parent cases dropped from 2.49 to 2.17. By March of the following year, the median grant amount had dropped by \$90, because TANF is a function of the family size and also the income of the family.

The figures collectively highlight the uniqueness of this natural experiment. Most studies of welfare reforms are hampered by the fact that individuals are rarely – if ever – removed from welfare *unexpectedly*. Rather, a small fraction of welfare recipients are timed out years after time limits are put in place, and families face the same program restrictions when they apply for benefits. Most welfare stays are short with few reaching the time limits (Ziliak, 2009). Indeed, studies of time limits rely on assumptions about welfare usage due to the fact that time limits are passed and only bind years later, if at all. More likely when recipients are removed, it is because they transfer to other welfare programs – receiving a comparable stream of income that would leave them unaffected – or else they are removed for cause due to failure to comply with program rules – complicating selection and any assessment of removal. We have found no other reforms that create quasi-random variation in welfare receipt from one period to the next for such a large number of people.

3 Data and Empirical Strategy

3.1 Data and Summary Statistics

This study uses administrative data from the Michigan Department of Health and Human Services (MDHHS). The dataset consists of the universe of welfare participants in the state of Michigan from 2008 – three years prior to the reform – through 2019. This includes information on the both the primary recipient and their family. The vast majority of cases are single parents with dependent children, specifically single women. Given the policy focus on the labor force participation of single mothers and the binding nature of the policies on this population, we focus our attention on this group of people.

We see, for each participant in each family, the TANF grant that family receives, along with other grants they receive like supplemental nutritional assistance (SNAP), their eligibility for Medicaid, along with their quarterly wages during this period, and whether they were on any welfare programming or not. This is critical as it allows us to model

labor force transitions both on, and off, welfare. We also observe a wealth of demographic characteristics about these families, the race/ethnicity of each family member, their gender, age, marital status, whether they are a migrant or veteran, their census tract of residence, and their final year of schooling.²³

We have additional variables that aid in our analysis. We observe counters for each participant that track how many months they have used benefits.²⁴ We can see overall applications to the TANF program, including acceptances and denials. We also can see when participants had cases closed due to noncompliance, reaching time limits, or other administrative reasons.

Table 1 displays summary statistics for the sample in bins based on their time on the program. There are evident patterns based on welfare participation. First, the grant is rising with time on the program. This is likely due to the family size increasing over the natural life cycle. However, longer-term welfare users are clearly more systematically disadvantaged than those newer to the program. Few families exhaust their benefits, self-selecting the remaining sample. The fraction of mothers working after 5 years on welfare is only 25%, compared to 33% for women with 2-4 years of welfare use. These women are older, have larger families, are overwhelmingly black, and 30% of them lack a high school diploma. There is also geographic dispersion in the sample. These families are overwhelmingly located in the Detroit metropolitan area. For studying welfare leavers due to the policy change, these statistics also illustrate the fact that it will be difficult to find a suitable control sample from among the welfare “remainers.” Those with shorter tenure on the program will likely not represent a suitable control group.

Most families never utilize all of the allotted time, or even come close. Taking the sample of people on welfare on or after 2008, the average utilization is 33 months, with the median use being 21 months, indicating several long-term beneficiaries skew the usage statistics. The 25th and 75th percentiles are 8 and 48 months, respectively. This includes the period when the welfare population, provided they had an eligible child under 18, was unbounded in their use of the program in the state. Nevertheless, as of September 2011, roughly 25% of all the single women on welfare in the state had been on the program for over 5 years. Figure 2 presents the details. The full distribution of welfare counters at the time of the policy is given in 2a where each line represents a discrete counter value in increments of 1. There are only a handful of families above 150 months on the program, and they are excluded here for legibility. Furthermore, for the purposes of analyzing the

²³Years of schooling and location are current as of their final appearance in the MDHHS system, so we cannot model education decisions endogenously. Moreover, we must assume that participants do not move throughout the time period studied.

²⁴This may differ from the number of months of TANF used due to exemptions.

Table 1: Summary Statistics By TANF Counter Value in September 2011

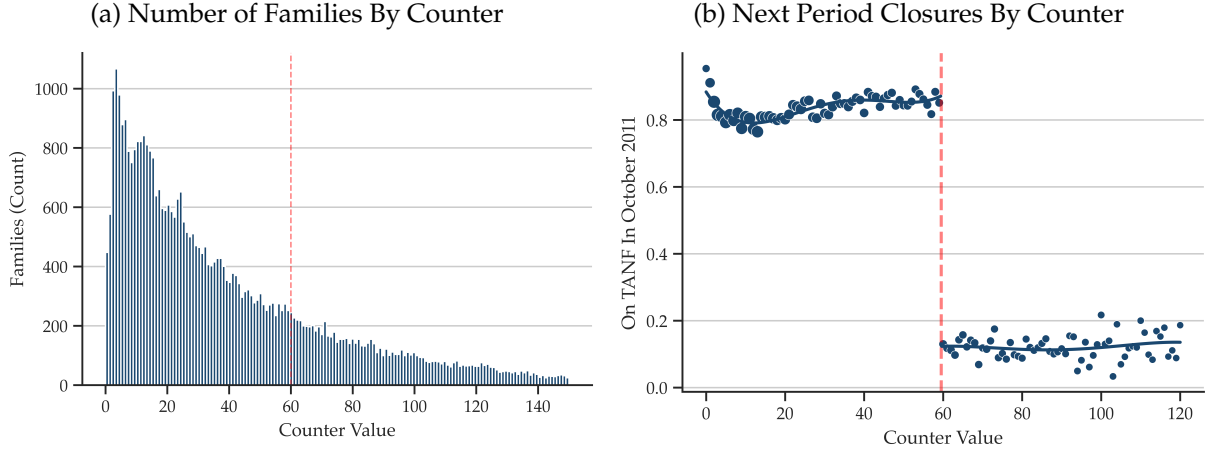
	0-1		2-4		5+	
	Mean	(SE)	Mean	(SE)	Mean	(SE)
Monthly Grant (\$US)	372.52	(1.17)	417.56	(1.47)	465.67	(1.77)
Quarterly Wages (\$US) ^a	716.64	(11.43)	704.44	(13.35)	443.3	(11.28)
Wages if Working (\$US) ^a	1999.68	(24.57)	2085.66	(30.26)	1729.53	(32.08)
Employed (%)	34.74	(0.36)	33.45	(0.41)	25.4	(0.44)
Age (Years)	27.64	(0.06)	30.28	(0.07)	35.47	(0.07)
Children (Count)	1.87	(0.01)	2.28	(0.01)	2.74	(0.01)
Age Youngest Child (Years)	3.72	(0.03)	5.19	(0.04)	7.53	(0.05)
Black (%)	52.69	(0.37)	67.17	(0.41)	80.82	(0.4)
Hispanic (%)	4.06	(0.15)	3.14	(0.15)	2.02	(0.14)
Schooling (Years) ^b	12.14	(0.01)	12.04	(0.01)	11.78	(0.01)
High School Diploma (%)	80.72	(0.29)	77.17	(0.36)	67.41	(0.48)
Beyond HS Education (%)	30.71	(0.34)	28.63	(0.39)	21.75	(0.42)
In Detroit (%)	36.58	(0.36)	46.27	(0.43)	59.26	(0.5)
On SNAP (%)	91.51	(0.21)	95.0	(0.19)	95.04	(0.22)
On Medicaid (%)	92.78	(0.19)	95.86	(0.17)	97.72	(0.15)
On Other Programs (%)	4.06	(0.15)	3.65	(0.16)	6.53	(0.25)
Observations	17,902		13,326		9,631	

Notes: This table shows summary statistics for single-mother households as of September 2011 based on their counter value at the time of the change in policy. The columns note the number of years their counter value has attained indexed from 0 (so a family between 0-1 would have used anywhere from 0 to 23 months). These numbers only include families enrolled in TANF at the time. (a) Quarterly wages averages over workers that have no earned income while wages if working removes non-working women and recomputes the mean. (b) A person is considered to have a high school diploma if they have completed 12 years of schooling. Of those who have some education beyond high school, this is most often partial completion of advanced schooling or the completion of a two-year degree.

experiment as a regression discontinuity, there is no manipulation of the running variable at the 60 month threshold.

Adjacent this figure in 2b is the “first-stage” of the policy as a regression discontinuity plot. We bin counter values using equal spacing mimicking variance and additionally allow the scatterplot bins to change size to reflect the distributional mass at each point (Calonico et al., 2017). We include flexible polynomials on either side of the cutoff for clarity. Conditional on being on the program in September 2011 (and true for any counterfactual period), there is about a 95% chance that you will be on welfare in the next month, if you have a counter value lower than 5 years (60 months). However, the policy had the effect of decreasing this probability to roughly 15% for those just above this cutoff. This drop of 80% reflects the fact that the survival rate of welfare users is not 100%

Figure 2: Cases and Closures as of September 2011



Notes: The first figure shows the distribution of cases by counter value in September 2011 where each line is one month. We stop the count at 150 for clarity, though there are a few families who fall about this top-coded value. Those to the right of the red dotted line were subject to immediate removal (barring exceptions) due to the policy. This is reflected in the second figure, which shows the drop in the probability of being on welfare in October 2011 conditional on being on welfare in September 2011. Here we bin counter values using equal spacing mimicking variance and additionally allow the points to change size to reflect the distributional mass at each binned value. The line is a flexible 4th-order polynomial.

from period-to-period and that the policy had exemptions for certain criteria related to hardship.

3.2 Empirical Strategy

We are interested in studying several different, though related, causal estimands. First, we wish to explore what happens when individuals suddenly lose cash benefits, since this serves as a test of income effects on labor supply models of low-income earners. Additionally, we wish to illuminate the direct effects of time limits themselves on welfare use and labor supply outcomes, since these results tell us about the optimal design of welfare programs and test assumptions of the forward looking rational behavior of welfare recipients.

Because of the nature of the reform, analyzing the policy is not straightforward. The policy creates treatment in the assignment of months, $M \in [0, 60)$, left on the welfare program – in other words, the effect of either being kicked off immediately, or relatedly, the effect of being assigned an integer number of months of remaining benefits. As M goes down, the treatment intensity increases since these individuals have fewer (or potentially no) welfare benefits to draw on for future use. We are interested in estimating the average treatment effects or quantile treatment effects of this assignment.

The fundamental problem we face is that, while the natural experiment created varia-

tion in the assignment of M , there are no completely untreated individuals. Any control group sampled from among this treated population at the time of the reform itself would come with the caveat that these people are still assigned a different value of M . In other words, in the case of immediate removal from welfare, there is a clear treatment group, but no counterfactual untreated control group. While individuals assigned larger values of M are treated less *intensely* they are still treated.²⁵

To circumvent this, our primary empirical specification uses the pre-period as a counterfactual for the treated sample by sampling individuals in the past with the same time on welfare. One should think of this as a modified difference-in-differences approach with respect to time. To be explicit, define Ω to be the treatment period (September 2011). We assume that an individual who has used D months of benefits at time period Ω – who is assigned $M = \max\{0, 60 - D\}$ months left on TANF due to the treatment – would have followed the same trend in outcome variable Y as an individual who had utilized D months of benefits in time period $\Omega - c$, where c indicates a particular number of months or quarters prior to the reform date Ω . This identifying concept – using variation in the assignment variable at different points in time to identify the treatment effect – is similar to an event study design.

Econometrically, and most simply, we are interested in identifying the coefficients in the following difference-in-differences regression model:

$$Y_{i\tilde{t}} = \beta_0 + \beta_1 W_{i\tilde{t}}^D + \lambda_i + \delta_{\tilde{t}} + u_{i\tilde{t}} \quad (2)$$

where i denotes the family, $W_{i\tilde{t}}^D$ indicates whether family i had counter value D on September 2011 (with 0 if they had the counter value D at some period in the past), $\delta_{\tilde{t}}$ are event-time fixed effects that soak up seasonal variation in outcomes, and $u_{i\tilde{t}}$ is an idiosyncratic error term assumed to be i.i.d. The only departure from the traditional difference-in-differences setup is the definition of \tilde{t} . Take t to be a given month in the sample, then:

$$\tilde{t} = \begin{cases} t - \Omega & \text{if treated}_i^D = 1 \\ t - \Omega - c & \text{if treated}_i^D = 0 \end{cases}$$

²⁵While theoretically individuals who are assigned, for example, $M = 40$ remaining months of TANF benefits may be treated less *intensely* than those immediately kicked off, they are still treated. Even if we were willing to make assumptions about the suitability of this group as a control population for the purposes of estimating a standard treatment effect model, as shown in table 1, these individuals are sufficiently different on observables to make selection a significant problem for identification. In other words, among potential within-treated control groups, the suitability of the population on observables is inversely related to the treatment intensity.

which means that if you are in the treated group, and the outcome is quarterly, then $\tilde{t} = [-1, 0, 1]$ corresponds to [2011Q2, 2011Q3, 2011Q4] and if we choose $c = 12$ months, then if you are in the control group \tilde{t} corresponds to [2010Q2, 2010Q3, 2010Q4].

Of course, c could be any arbitrary past month of benefits. This highlights the drawback of this particular normalization: the closer we make the previous date of comparison, the more similar are the prevailing wage and economic conditions (better counterfactual), however the shorter the potential post-period for estimation, since the control group itself is treated on September 2011. Our baseline specification uses September 2010 as the comparison period of interest. We show robustness to all available time periods in the appendix (see section 5).

In the analysis that follows, we also specify event study regressions to explore more complicated dynamics and treatments that vary across individuals over time – as opposed to events that occur with the same timing for all participants. Unless otherwise specified, we follow [Schmidheiny and Siegloch \(2020\)](#) in defining the bins and event study specification, and we estimate the following equation

$$y_{it} = \alpha + \sum_{j=\underline{j}, j \neq -1}^{\bar{j}} \beta_j b_{it}^j + \mu_i + \delta_t + \epsilon_{it} \quad (3)$$

where b_{it}^j are the binned event indicators of the form

$$b_{it}^j = \begin{cases} \mathbb{1}\{t \leq e_i + j\} & \text{if } j = \underline{j} \\ \mathbb{1}\{t = e_i + j\} & \text{if } \underline{j} < j < \bar{j} \\ \mathbb{1}\{t \geq e_i + j\} & \text{if } j = \bar{j} \end{cases}$$

with e_i denoting the event time operator. *Events* in this case could include sanctions due to noncompliance or timing out of welfare. y_{it} is the outcome for individual i at quarter t , μ_i are individual specific fixed effects, and δ_t are time fixed effects. As specified in the equation, we normalize event time period -1 to 0 .

Given the recent literature that notes the inconsistency of traditional difference-in-differences estimators when there are multiple treated cohorts – due to the potential of negative weights in the resulting estimator – we follow [Gardner \(2021\)](#) in estimating the regression in an event study context.²⁶ Following this design, we estimate the equation in two steps to remove the cohort and period fixed effects from the variable of interest

²⁶This is one of many potential corrections that all aim to achieve similar results; see [de Chaisemartin and D’Haultfœuille \(2020\)](#) or [Goodman-Bacon \(2021\)](#) for a more detailed discussion.

prior to normal estimation in order to achieve consistency. Specifically, we first estimate the model

$$Y_{gpit} = \lambda_g + \gamma_p + \epsilon_{git}$$

where Y_{gpit} is the same outcome Y_{it} in equation 3 except now we note its dependence of the treatment cohort g and period cohort p , with accompanying fixed effects λ_g and γ_p . Here we have divided the individuals and time into treatment groups $g \in \{0, \dots, G\}$ and periods $p \in \{0, 1, \dots, P\}$ defined by the adoption of the treatment, where members of treatment group 0 are always untreated, 1 are treated in period 1, 1 and 2 are treated in period 2, and so on. We estimate this equation using only the sample for which the treatment indicator equals 0, retaining the estimated coefficients, and then in the second step use $Y_{gpit} - \hat{\lambda}_g + \hat{\gamma}_t$ in place of Y_{it} in the event study regression. Due to the parallel trends assumption, this procedure identifies the effect of the treatment, even when the adoption and average effects of the treatment are heterogenous with respect to groups and periods.

In robustness checks and in further analysis of the primary treatment effects, we make use of a regression discontinuity design and unconditional quantile regressions in the style of [Firpo et al. \(2009\)](#). These are discussed in more detail in the relevant subsections and in the appendix.

4 Main Results

4.1 Removal From Cash Welfare

4.1.1 Average Results

The immediate impact of the policy change was to suddenly remove thousands of families from cash welfare. Our baseline specification uses equation 2 to estimate various outcomes for this group of families by taking the *treated* group as those families with between 5-6 years on welfare at the time of the policy change, and the *control* group as those families who had between 5-6 years on welfare one year in the past.²⁷ The pre- and post-periods for each respective group are one year before and after the (counterfactual) event. As explained in section 3.1, we focus on female-headed single-parent households as this group of families form the majority of TANF cases and are a uniform population

²⁷Theoretically, we could include all affected families with counter values about 5 years; however, this sample limitation prevents overlap in the treated and control groups. Results are unaffected by this restriction.

Table 3: Regression Results of Welfare Loss

	TANF	Employed	Wages	Budget	Medicaid	SNAP
Estimate	-0.41** (0.01)	0.04** (0.01)	476.62** (147.27)	-1864.40** (161.43)	-0.02** (0.01)	0.01 (0.01)
Constant	0.76** (0.00)	0.25** (0.00)	2077.14** (41.67)	14835.80** (42.75)	0.96** (0.00)	0.92** (0.00)
F-Statistic	1376.87	18.51	10.47	133.39	8.34	1.79
N	121400	43704	43704	121400	121400	121400

The results in this table come from estimation of separate two-way fixed effects models on each column variable. SNAP, TANF, Medicaid, and Employment are all binary variables and thus estimates are interpreted as probabilities. Wages and Budget are in annualized US dollars. Standard errors in parentheses are clustered by family. **, *, + indicate p-value less than .01, .05, .1, respectively.

of interest.

Table 3 presents the main results. Each column gives the estimate on a separate outcome of interest. The first column represents the change in the probability that the participant is enrolled in TANF (relative to the control group) in the year after treatment. Compared to the control group mean of 77 p.p, enrollment in TANF is approximately 42 p.p. lower among the treated group. This *first stage* estimate differs from the apparent 80 p.p. drop seen figure 2b for three reasons. First, individuals naturally cycle off of TANF; indeed as figure 2b shows, approximately 10% of the sample does not continue on TANF from one period to the next, even at high past usage, either due to choice or failing to meet program requirements. Second, the court case temporarily reinstated welfare payments for individuals over the post-period, which means that the probability that individuals are on TANF rises after February 2012 for the treated group. Finally, approximately 20% of the individuals are exempted from the policy due to hardship criteria. As a result, any final estimates of the effects of the policy on labor supply should be seen as lower bounds of the true effect.

The following two columns present the primary labor market outcomes: the effect of the policy on the probability that the individual is working and the effect of the treatment on the wages of these families. Both of these estimates confirm that on average, the policy induces some families to work. The policy results in an increase of formal employment of roughly 4 percentage points over the control group mean of 25% employment, implying a nearly 20 percent increase in formal labor market participation by these families. In column three, we see that the policy has increased the average (annualized) earnings of these families by roughly \$500, again representing a 25% increase over the control group mean.

However, it is important to note that the majority of families still remain out of the

formal labor market. As pointed out in a growing literature on “disconnected mothers”, an alarming number of former welfare recipients have neither transitioned to formal employment nor to other forms of social assistance when exiting cash welfare programs (Blank, 2007; Blank and Kovak, 2009; Turner et al., 2006). This appears to be the case for the majority of participants removed from Michigan’s TANF program. While the numbers above suggest that removing families from welfare can increase formal employment, it also shows that the vast majority remain unemployed, meaning that these poor families become even poorer. We can see this in column four. Using our administrative data, we create a monthly budget of these families by adding up their quarterly wages (divided by three and allocated to each month), their monthly SNAP benefit (if enrolled), and any money they receive from the TANF program – these programs combined represent the majority of formal earnings for this population – and calculate the treatment effect of the policy on this budget (annualized for comparability). On average, these families are almost \$2000 a year poorer after being removed from TANF. Relative to the control group mean of \$15,000, this is a significant income penalty, particularly since these individuals are supporting children.

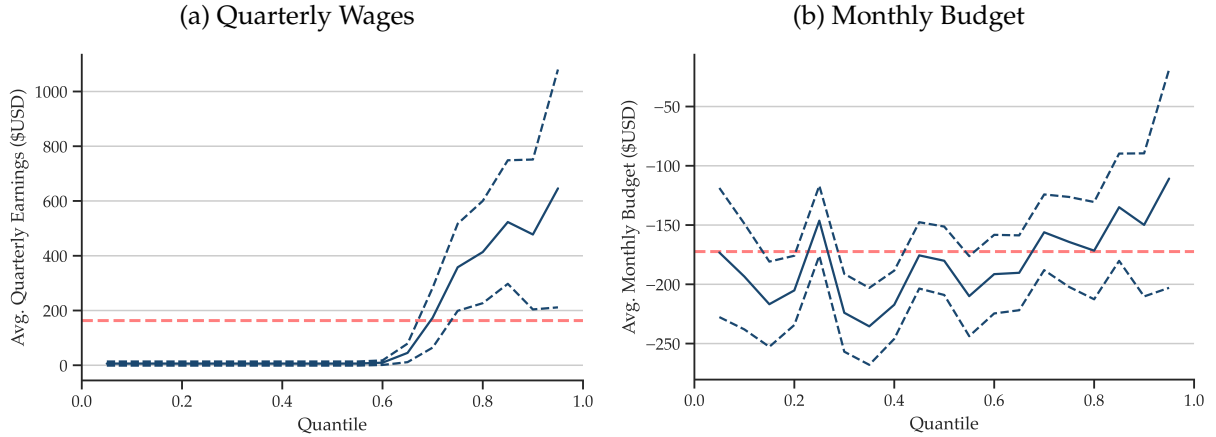
Finally, the remaining columns show the flow into two programs that are vital to poor families, Medicaid – the primary medical program for poor families – and the supplemental nutritional assistance program (SNAP) – a program which subsidizes grocery purchases for low income households. Participation in both of these programs is near universal for this sample. We see that the policy reform leaves SNAP unchanged – unsurprising since low-wage workers are also eligible for SNAP – but reduces medicaid enrollment. Enrollment in these programs is handled slightly differently, and changes to the enrollment process combined with losing touch with welfare offices may explain this discrepancy.

4.1.2 Unconditional Quantile Results

As shown in table 3, the policy has positive extensive margin effects on average, but most families do not work. Buried within this treatment effect then is heterogeneity in responses to the policy, a point noted in existing welfare reform literature (Bitler et al., 2006). We explore this heterogeneity in two ways, starting by estimating unconditional quantile regressions of the difference-in-differences estimate.

To estimate quantile treatment effects, we follow the two-step methodology proposed in Firpo et al. (2009). In the first step – for each quantile q analyzed – we use a re-centered influence function to transform the endogenous variable. In the second step, we regress

Figure 3: Unconditional Quantile Treatment Effects



Notes: These figures show the results of estimating unconditional quantile regressions following [Firpo et al. \(2009\)](#) for quarterly wages and the monthly budget (SNAP + TANF + wages) as described in the main text. The red dashed line shows the rescaled result found in table 3. We plot estimates at each 5th quantile of the distribution with 95% confidence intervals in dashed lines around each point estimate in blue.

this transformed variable on the same regressors as before, where now coefficients can be interpreted as the marginal effects on the q^{th} unconditional quantile of the distribution of the outcome variable. To neatly display results for the entire distribution, we run the regressions at every fifth quantile between the 5th and 95th quantile of the outcome variable distribution. Results for each quantile – along with the average OLS estimate (rescaled) from table 3 (in red) – are shown for the two monetary outcomes in figure 3. Each point, along with 95% confidence intervals, is the difference-in-differences coefficient of interest at that quantile.

Looking first at the leftmost figure, the treatment effect on quarterly wages is 0 for the majority of the distribution. Beginning at around the 70th percentile, the policy is statistically significant and begins to increase earnings above the average OLS estimate. By the 95th percentile, the effect is nearly \$650 a quarter, a sizable increase. However, even at this quantile, it is clear that families are worse off financially. Informally, the average TANF grant is larger than \$300 per month; therefore, an increase in wages of \$600 a quarter (\$200 a month) will still leave families worse off financially, putting aside increased costs of child care or commuting costs to employment.

More formally, we look at the effect of welfare removal on the distribution of the constructed monthly budget in the rightmost figure. With point estimates ranging from -\$250 to -\$100, at every point along the distribution, the estimates are negative, and rarely are they different than the average point estimate of the decrease in monthly budget as calculated prior. In other words, welfare removal makes families more financially vulnerable

along the entire distribution grant receipt.

4.1.3 Heterogeneity in Treatment

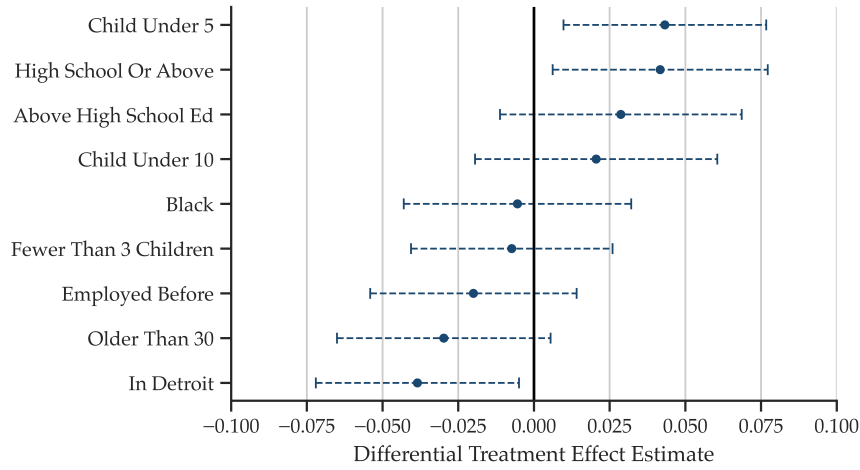
In addition to quantile treatment effects, we are interested in other heterogeneity of the policy. Ex-ante, we would expect that individuals who are older or have lower observed education levels would be less able to find work after losing welfare. More generally, welfare removal likely affects individuals differently depending on if they have young kids, are older, have not completed high school, or are from historically disadvantaged racial groups. To examine this, we simplify the analysis by pre-defining policy relevant covariates and estimating the following triple difference equation to test each potential source of heterogeneity:

$$Y_{it} = \beta_0 + \beta_1 \text{after}_t + \beta_2 \text{treated}_i + \beta_3 G_i + \beta_4 (\text{after}_t \times \text{treated}_i) + \beta_5 (\text{after}_t \times G_i) + \beta_6 (\text{treated}_i \times G_i) + \beta_7 (\text{after}_t \times \text{treated}_i \times G_i) + u_{it} \quad (4)$$

where each covariate tested is a binary variable. The terms after_t and treated_i denote whether the period is one year after or one year before the normalized event time of the policy and whether the unit is in the treated group, respectively. G_i is a covariate of interest, such as those described above. A full list is provided below. The coefficient of interest is β_7 , which estimates the differential average impact of the treatment on a subgroup with a particular value in the covariate space “turned on.”

To target the analysis further, we focus on the binary decision of whether the (former) participant is employed as a result of the treatment. We estimate the triple difference for the following binary covariates: head of household is black, they are above the age of 30, their youngest child is under the age of 5 (10), they were employed in the year prior to treatment, their highest completed grade level is equal to (below) high school graduation, and whether they have 2 (or fewer) children. The age of 30 and 2 or fewer kids corresponds (roughly) to the median age of women in the sample and the median number of children these women have, while the ages of 5 and 10 correspond (roughly) to the median and 75th percentile of ages of children in the sample, respectively. Figure 4 presents the estimated coefficients for each of these interactions. Each term on the y -axis is the variable estimated and each point and 95% confidence interval corresponds to the estimate of β_7 . Due to the inclusion all interaction terms, each estimate represents the difference from the mean treatment effect of the excluded group. For example, the coefficient on black represents the difference between treatment values of black welfare

Figure 4: Heterogeneous Treatment Effects For Employment



Notes: This graph shows the estimates of β_7 in equation 4 separately for each binary outcome of interest as defined in the main text. 95% confidence intervals are shown with dashed lines. All estimates are thus in relation to the primary treatment effect for the employment increase.

participants as compared to white ones.

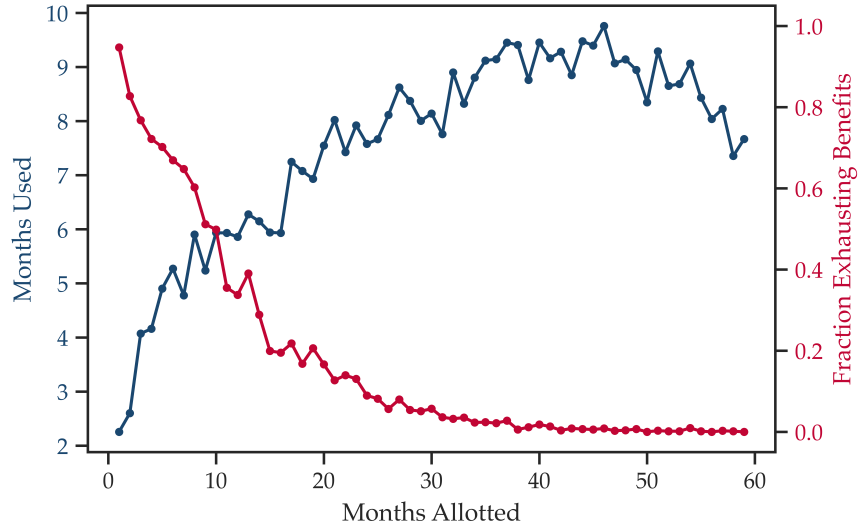
Those with high school or above education or those with youngest children – most likely to be younger and excluded from welfare time limits, and as a result more likely to be temporarily out of the labor force – have positive and statistically significant interaction terms. These individuals are almost 5 p.p. more likely to be employed in the post period relative to those without a high school education and those with older children. This indicates that human capital investments mitigate these negative income shocks. Conversely, those living in Detroit are differentially less likely to be employed. This is a sizable share of the welfare population in Michigan, and reflect the fact that high concentrations of urban poverty can causally raise barriers to employment.

4.2 Time Limits

While the policy had the effect of suddenly removing thousands from welfare rolls, it simultaneously applied time limits to the remaining families. Prior to the reform, families expected to be able to use the TANF program until their youngest child turned 18; following the reform, their remaining time allotment was determined by their prior use in combination with the age of their youngest child.

Figure 5 illustrates how many months beneficiaries use on average by allotment and how many ultimately reach benefit exhaustion. The x -axis is determined as the new 5-year time limit minus prior usage at the time of the reform. We show two different y -axes.

Figure 5: Mean Usage and Exhaustion Percentage By Time Limit Allotment



Notes: The axes for this figure represent the cumulative months used of TANF (y-axis 1) over the period from October 2011 through December 2019 – the last date for which we have data – or the fraction of total assigned months used (y-axis 2) over the amount of months assigned remaining due to the policy on the x-axis, M , where $M = \max\{0, 60 - D\}$, with D the number of months of TANF used prior to October 2011 by the given family.

The blue line shows cumulative welfare usage in months from October 2011 through December 2019 (excluding the brief period in 2012 when the court case artificially reinstated benefits for most participants).²⁸ While months utilized is rising with the total allotment sharply in relative terms, this relationship is not steep in absolute magnitudes. Families allotted 20 months use 7 months on average over the next 8 years, nearly the same as those allotted 40 months, who use 9 months, on average. The relationship plateaus at around 40 months and seems negatively related after this, likely due to the high exit rate of families on welfare (even prior to the reform) among program entrants.

We plot the average fraction of families who have exhausted their benefits by December 2019 as a function of the number of months they were assigned in 2011 in red.²⁹ The figure shows a clear downward trend in benefit exhaustion as a function of the number of months assigned. The majority of families assigned fewer than 10 months ultimately reach the time limit. However, the speed at which the drop-off in benefit exhaustion

²⁸The resulting patterns look nearly identical even if including these months. Rather than benefit usage kinking at 0 for those over the 60 month time limit, the results kink at approximately 6 months of use, meaning that the reinstatement of benefits uniformly gave families 6 months more usage than they otherwise should have expected. The court case only affected people who had used enough months at the beginning (or throughout) the court case period, that their benefits should have been exhausted due to the time limits.

²⁹Due to this graph illustrating how many reach the time limit (and could include people who go *over* the limit), we do not exclude the court case period. So this is an upper bound of the number that exhaust benefits.

happens is surprising. For those assigned 2 years of benefits, fewer than 20% of families exhaust all their benefits; after 3 years, this number is nearly 0% where it stays for the remainder of the distribution.³⁰

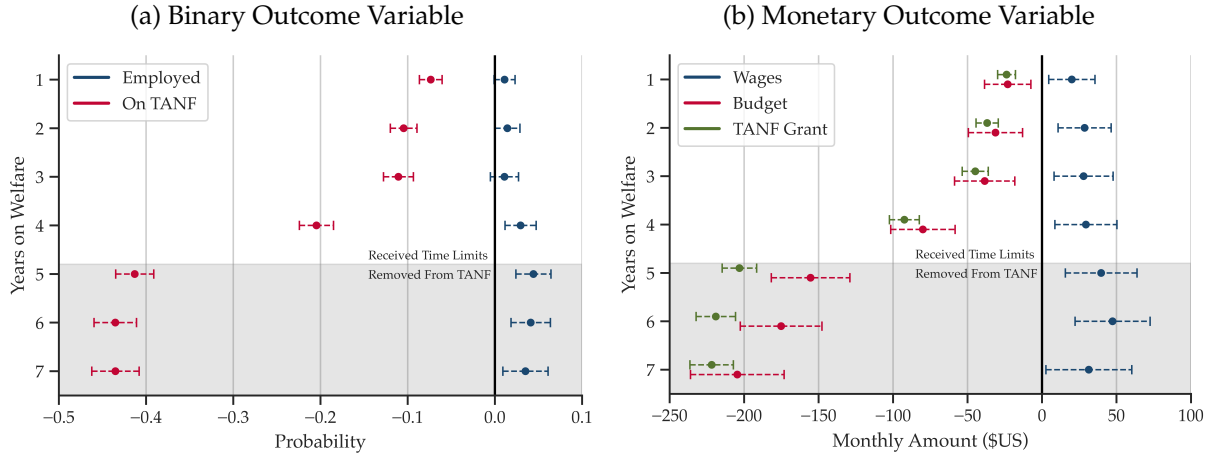
4.2.1 Overall Effects of Time Limit Assignment

To see how families react to the imposition of time limits, we begin by analyzing the overall effect of time limits on welfare use and labor supply using the arguments laid out in section 3.2, which apply to time-limited families as well. Specifically, we estimate equation 2 for individuals with 1-7 years accrued on welfare instead of just those immediately over the threshold as in table 3. For those individuals who had not reached the 5-year time limit, the results should be interpreted as the effect of being assigned $M = \max \{1, 60 - D\}$ months of benefits conditional on having already used D months, compared to the counterfactual world where one had used D months and could use any amount of benefits until their youngest child turned 18. For those above the 5 year limit, the interpretation is the same as in section 3.2, except that doing the analysis in this way enables comparison between those who were unexpectedly removed from welfare and had 5, 6, or 7 years on the program at the time of treatment.

The results are summarized in figure 6. In each subfigure, the y -axis is the numbers of years the family has been on the program. Results should be interpreted with the (imposed) five-year time limit in mind. For example, someone who has been on the TANF program with y -axis value 4 has been assigned < 12 months remaining on welfare. The x -axis is either a probability – from the linear probability model specification in the case of a binary dependent variable – or a dollar amount – transformed to be in terms of the monthly budget of the family. Each point is the point estimate of β_1 from the two-way fixed effects model 2 with accompanying 95% confidence intervals, clustered by family. To ease interpretation, the figures are highlighted to show the region at which the treatment would change from being assigned a time limit (no shading) to being removed from TANF entirely (shaded). Even though the treatment is the same for everyone above 5 years, we still show the treatment separately for years 5-7 as robustness and as a way to explore whether these welfare long-haul users have the same treatment effects as those

³⁰In appendix section A.3 we provide further analysis of the variation by comparing the distributions of welfare use by years assigned and also the fraction of remaining time used. Overall, all those with access to more than 3 years of benefits seem to use similar levels of benefits. Those assigned fewer months, tend to exhaust benefits. We note that time limits seem to create heterogeneity in welfare usage, but the puzzle is whether this heterogeneity is due to the mechanical effect of losing benefits after a period of time, observable differences in the pre-existing population at each time limit assignment, or behavioral differences in banking that occur due to the change in the eligibility horizon and the incentive to bank benefits that families face with the institution of limits on use.

Figure 6: Difference-in-Differences Results At Various Counter Values



Notes: Each point in the above graph represents an estimate of β_1 from equation 2 for a given counter value (the y -axis) in the year after the treatment, alongside 95% confidence intervals. The shaded region represents the point at which the treated individuals go from being treated with a time-limit assignment to being treated with welfare removal. The control group is always the individuals with the same counter value one year prior who were living under the assumption that their welfare was not time limited and could be used until their child turned 18. In order to have everything on the same axes for comparability, we divide quarterly wages and the budget by 3 to compare it with the TANF grant the family received in the second subfigure.

just over the 5 year limit. This may be the case if, due to longer absence from the labor market, these families are less likely to be gainfully employed even after losing welfare.

In figure 6a, the point estimates show that the effect of time limits grows with years on welfare. That is, individuals assigned fewer months of benefits have strong labor market responses, despite the fact that these families have also been on welfare the longest, and therefore are ex-ante the least likely to find stable employment. The treatment is nonzero for those well below the threshold for being removed; this is consistent with the findings of Grogger and Michalopoulos (2003) who show that those women far away from the cutoff face more incentive to bank benefits as insurance for negative wage draws in the future. Our results suggest that the policy change causes an additional 10% of these families to leave welfare early, relative to a non-time limited counterfactual.

Additionally, the treatment effect is virtually the same for each family with more than five years accumulated on the program – who all are about 50% less likely to be on welfare the following year – and those who have been on for three or fewer years – who are 10% less likely to be on welfare – respectively. For those with four years on the program, this group has both treatments in the post-period. They are assigned time limits between 1-11 months left on welfare, and many of them subsequently are removed when they reach these limits. Therefore, the probability that they are off welfare in the following year approaches the probability of those with 5 or more years on the program.

In the same figure, we see the effect of the treatment on the probability of having positive earnings in the year following the policy. This probability also rises monotonically with time on welfare; the closer the end of the eligibility horizon, the more likely the family is to find work in the year after, relative to the control group. While this is consistent with economic theory about labor supply, it is notable given that these women have been out of the labor force longer, by definition. In figure 6b, in the same color, this is reflected in a monotonic increase in monthly wages with time on the welfare program. Those with 1-3 years of welfare used – who are assigned 4-2 years left of benefits, respectively – are making approximately \$20 more on average than the control group per month. This rises to \$50 on average for those kicked off. This represents an increase in the probability of gainful employment and wages of about 8% of for those who receive a time limit as a treatment, compared to 25% in each of these labor market indicators for those who are removed completely (as described in the last section). Not only are these estimates consistent with what we would expect from forward looking decision makers and from individuals who faced a large negative income shock, but the numbers are relatively large.

However, figure 6b also shows the effect of the policy on the monthly budget of these households. While some women start working and on average these families do not make this back in wages, the net effect of the policy is negative on their overall wealth. As expected, this negative effect is minor for those well below the eligibility threshold, since they can always choose to remain on welfare if they do not have a wage draw above their reservation wage. The red point estimates show the effect on the total budget of these families – the total of their welfare, SNAP, and TANF benefits in each month. Consistent again with the intuition of Grogger and Michalopoulos (2003), we see that families treated with the imposition of a time limit have a total monthly budget about \$50 lower relative to the no-time-limit control group. Although this is statistically significant, it is small in relative terms. This negative amount rises to four times that for those who were kicked off, showing that the policy had the effect of making some of the poorest families in Michigan poorer, rather than lifting them out of poverty.

4.2.2 The Marginal Value of a Month of Welfare Eligibility

As noted most famously in Grogger and Michalopoulos (2003), the imposition of time limits affects the decision making behavior of individuals due to the fact that forward-looking utility-maximizing households use welfare as a form of labor market insurance. Given two individuals with the same time horizon with which to use benefits (the same

age of their youngest child), we expect that mothers with fewer months of benefits remaining will be more cautious with their use.

To illustrate the point, consider two mothers who both have already used 3 years of benefits, and assume one has a youngest child of age 7 and the other a child of 15. Both such families are *allocated* 2 remaining years of benefits due to the policy. However, the former has 11 years to utilize those benefits, while the latter must use those 24 months of eligibility within the next 36 months, or else lose them forever. Due to the use-it-or-lose-it nature of welfare benefits, in addition to the impulse to conserve benefits for use in the future (for potential negative wage draws), we expect that the former family will exit welfare earlier than the latter family.³¹ Thus the same exact time limit can have heterogeneous impacts on families depending on the age of the youngest child.

The above discussion, alongside the extant literature, illustrates that time limits affect welfare use – directly – and thus labor supply decisions – indirectly. A natural question to ask then is what is the elasticity of labor supply or welfare use with respect to an additional month of welfare eligibility? Relatedly, do women with the youngest children at the time the policy is implemented leave welfare earlier than those with older children, even though they have the same time limit?

To explore this empirically, we run regressions of employment and wages in the three years following the policy – and welfare use, SNAP use, and Medicaid enrollment in the eight years following the policy – on the lost (potential) months of welfare benefits due to the policy.³² Here we define the lost months of welfare benefits as:

$$\text{Lost Months} = 18 \times 12 - A - \max \{0, 60 - D\}$$

where A is the age of the youngest child in months and $\max \{0, 60 - D\}$ is the amount of months of benefits allocated to a family who has used D months of benefits prior to the imposition of the policy. This definition then is the *maximum* number of months of potential welfare benefits that the mother has lost on the date the policy comes into place assuming they were to remain on welfare for the full tenure of their eligibility. Referring again to the example of above, the family with a 7 year old youngest child has lost 108 months of potential welfare use, while the family with a 15 year old youngest child has only lost 12 months.

To isolate the effect of the loss in expected eligibility on the outcomes listed above,

³¹This argument follows from the reasoning of [Grogger and Michalopoulos \(2003\)](#), provided that families are forward looking and rational.

³²Due to our data use agreement, this is the maximum horizon over which we have access to data on employment and wages.

Table 4: The Marginal Value of a Month of Welfare Eligibility

		On TANF	Employed	Wages	SNAP	Medicaid
All Families	Elasticity	−0.18** (0.03)	0.15** (0.03)	0.31** (0.07)	−0.01 (0.03)	−0.04* (0.02)
	Observations	16798	14843	14843	17419	17553
	Adj. R-squared	0.19	0.16	0.17	0.08	0.05
	F-statistic	452.19	1706.34	1424.08	82.92	92.86
Young Child	Elasticity	−0.90** (0.23)	0.85** (0.21)	1.90** (0.39)	−0.53** (0.16)	−0.25 (0.16)
	Observations	9492	8452	8452	9809	9887
	Adj. R-squared	0.21	0.16	0.18	0.09	0.06
	F-statistic	180.82	1118.0	915.64	91.61	71.83
Older Child	Elasticity	−0.12** (0.04)	0.06 (0.05)	0.13 (0.11)	0.05 (0.03)	−0.00 (0.03)
	Observations	7306	6391	6391	7610	7666
	Adj. R-squared	0.16	0.16	0.17	0.07	0.05
	F-statistic	130.41	740.65	656.2	41.42	41.9

Notes: This table shows the results of the absorbing least squares regression with the explanatory variable the log of months lost on the date of the policy. The outcome variables are, respectively (in log terms), cumulative TANF use for the subsequent 8 years, cumulative quarters employed and total quarterly wages for the subsequent 3 years, and SNAP and Medicaid enrollment for the subsequent 8 years after the policy. We use fixed effects to isolate the effect due to the lost months of eligibility. The fixed effects included are months of TANF assigned (60 minus months used), number of children, age in years, high school graduation status, race, and county of residence. All regressions control for employment, wages, and TANF use in the year leading up to the reform. The sample includes all women between the ages of 18 and 30. This restriction on ages accounts for the majority of the sample and isolates the women with the lowest barriers to employment and those with children of approximately the same age. The age cutoff for young child is 2 years of age, which is approximately the median child age. Standard errors, clustered by months of TANF use, are in parenthesis below each coefficient. ** Indicates p -value is less than .01, * Indicates p -value is less than .05, + Indicates p -value is less than .1.

we shut down all other potential sources of variation between the families, using absorbing fixed effects. Our baseline specification saturates the model with fixed effects for the time-limit assigned, number of children, age in years of the mother, county of residence, and education. We focus on women under the age of 30 at the time of the policy (who after 10 years will still be of working age). We include continuous controls for earnings, employment, and welfare participation in the year before September 2011. We cluster standard errors by the time limit assignment. In the appendix we provide results where we relax some of these restrictions; as a summary, we find that only the number of children and time limit assigned are critical absorbing variables, since these determine the welfare assignment rule. The other variables merely add precision to the estimates.

Results are shown in the first row of table 4. The (in)dependent variables are all in

log terms, so the coefficients should be interpreted as elasticities. We find that a 1% percent decrease in the welfare eligibility window leads to a 0.18% decrease in welfare use over the following eight years. It subsequently leads to a 0.15% increase in employment and a 0.31% increase in wages over the three years following the policy. Unsurprisingly, the policy has no effect on SNAP due to the fact that working families continue to receive SNAP benefits far past the welfare eligibility threshold. There appears to be a small effect on Medicaid enrollment; this may be due to participants gaining healthcare coverage through employment. Regardless, it is not large in absolute terms. In the appendix, we run the same results but in levels. Each additional lost month of eligibility is associated with a 0.02 fewer months of welfare use over the following 8 years. Considering a mean number of welfare eligible months lost of 150 with a standard deviation of 25, moving from the 5th to the 95 percentile of potential welfare eligible months lost implies 2 months fewer use in the post period. In terms of labor market outcomes, each additional month lost implies a .07% increase in employment and a \$43 increase in earnings over the following three years per month lost. While not large, it is important to keep in mind that these effects are merely due to the shortening of the *eligibility horizon*, as opposed to losing actual months of potential use.

To further examine the implications of the model of time limits first put forth in [Grogger and Michalopoulos \(2003\)](#), we break the sample at the median child age of youngest child – 2 years old – and run the same regression on these subsamples. Recall that the model predicts that women further from the end of their eligibility ending – those with the youngest children – will face the largest incentive to bank benefits. Consistent with this idea, we find that the perceived loss in eligibility window has the strongest effects on those women with the youngest children. For them, a one percent increase in months of eligibility lost due to the policy is associated with a 0.9% decrease in welfare use over the following years. Similarly, the employment and wage response are strongest for this group. For women with children older than 2, the estimates are in the same direction as those with younger children, but the magnitudes are far smaller and not statistically significant. In the appendix, we repeat the analysis of table 4 for those assigned more (less) than 3 years on benefits – roughly corresponding to the visual differences in the distributions we observed above – and find much stronger effects for those assigned more years left on benefits. Overall, this analysis is consistent with prior theory, and provides further evidence that time limits affect welfare use even far from benefit expiration, since participants have incentives to bank benefits to use at later dates.

4.3 Work Sanctions

An element of the welfare eligibility rules that has changed since PRWORA is that participants face strict sanctions for failing to comply with program requirements, specifically work requirements. These penalties were meant to encourage labor force participation and also raise the implicit cost of welfare receipt by imposing an administrative burden on beneficiaries. Despite the importance of these requirements on the welfare literature and in structural models of welfare take-up ([Chan, 2013](#); [Gray et al., 2019](#); [Meara and Frank, 2006](#); [Swann, 2005](#)), the actual empirical effects of these policies on labor supply are not well understood, largely because of a lack of available administrative data that contain information about individuals both on and off welfare alongside records of any sanctions imposed.

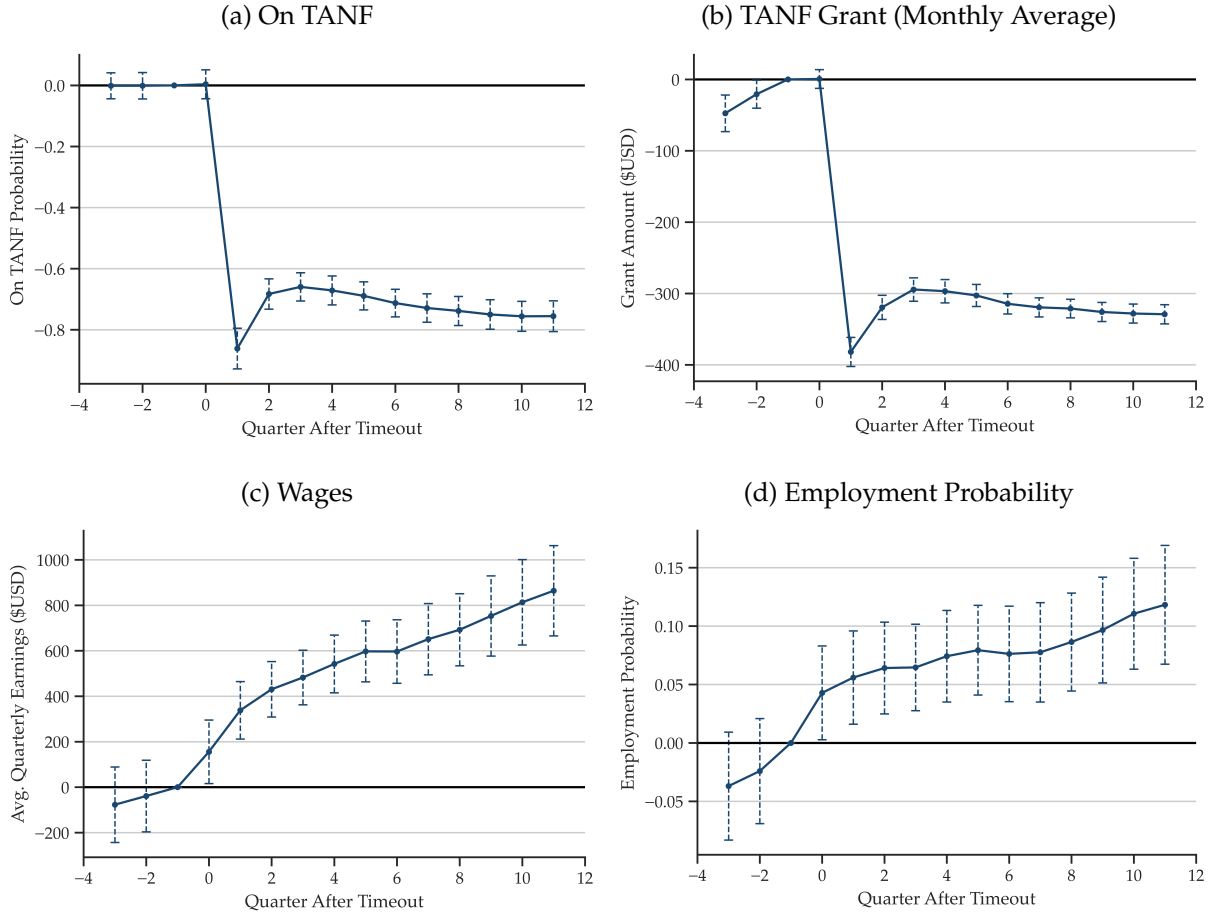
Furthermore, analysis of work requirements has largely overlooked the fact that states vary in what counts as work for the purposes of satisfying program rules and how strict they are about enforcement. In particular, it is no doubt the case that discretion and random audit potential influence compliance with program rules; participants may vary their compliance based on the likelihood of being caught, and this creates random variation in the timing of sanctions.

Michigan, like other states, implements a three-strike rule for failing to meet work requirements, with each subsequent violation raising the severity of the penalty. The first violation results in a 3 month ban from TANF (for the whole family), the second a 6 month ban, and the final a lifetime ban. The time limit continues to count for individuals while they serve the ban. Recipients are not sanctioned each time they break a work rule; rather, the rule must be broken without *cause* – unplanned events, illness, or structural factors that prevent the individual from completing their work requirement.

To shed some light on the dynamics surrounding these sanctions, we again use the data we have about case closures. Of the 172,061 families in our sample over the 12 years for which we have data, we observe 51,693 first time sanctions, 15,416 second time sanctions, and 4,097 third time sanctions. Sanctions are not randomly applied, nor is there any way for us to make a strictly causal claim about the data: we do not observe the petitions made about the cause of the work rule broken nor do we have information on the decision maker to model their decisions. However, the goal is to answer a basic empirical question: what happens to formal labor market outcomes after individuals receive a welfare sanction, as compared to those who do not?

Since we have much more data for first time offenses, and these sanctions are the most temporary from the perspective of the family receiving aid, we focus our attention on

Figure 7: Event Study Results For Work Requirement Sanction



Notes: This figure shows the event study coefficients from the regressions for the welfare sanctions as defined in the main text. Welfare sanctions are 3 month suspensions of welfare. We constrain the sample to only those people who ever received a sanction and additionally ensure that individuals were on TANF for at least 6 months prior to being sanctioned.

these events. We eliminate all individuals who never failed to meet the work requirement, so that we are left with a sample that is likely more balanced.³³ We use equation 3 to estimate the effect of the sanction on the family, relying on the differential timing of events for identification. The only additional sample restriction is that we require that individuals are on TANF for at least 6 months prior to being sanctioned in order to be counted as a valid event to ensure that the individual is fully aware of welfare compliance rules and had been in compliance (at least formally) up until that point in time.

Figure 7 shows the results for the probability the family is on TANF, the grant amount, their wages, and their work probability. The binned end-points are removed for clarity,

³³Results are similar when including the never treated individuals, however interpretation is complicated, and the pre-treatment parallel trends do not always hold for all variables. We omit them for this reason, to focus on the population that (at some point) received a sanction.

and we show the remaining dummies all in relation to the quarter before the sanction. As is evident from the first figure, the families are technically on TANF in the quarter they receive the sanction, however, they may receive the sanction at any point in the quarter, which is why the period 0 effect is smaller than others for the labor force participation, for instance.

The first thing to note is that the sanctions only lead to a decrease of roughly 80 p.p. in terms of TANF use. That is because the control group has individuals naturally cycle off; additionally, since the event is defined quarterly, some participants are put back on by the time the average is taken for period 0 in TANF use. Another interesting thing to note is that some families return; however, this is much smaller than the original drop, suggesting that work requirements function as a sufficiently high deterrence from families returning to TANF use. Rather than going back to zero difference, in the second quarter after the sanction, there is an increase of roughly 15 p.p. in welfare use, but then this too phases out over time back to 80 p.p. difference relative to the control group baseline.

After being removed from TANF, some families seem to return to the labor force. There is a roughly 5 p.p. increase in formal labor force participation that increases over time to 11 p.p. These families end up making roughly \$800 on average per quarter more than the control group three years after the treatment. However, \$800 per quarter, or \$2400 a year does not come close to restoring the lost income from the TANF grant itself. This is thus consistent with the earlier evidence about quasi-random removal and the effects of time limits. Some families do work, but most do not. The interesting thing to note about this sample though is that most (if not all) of these families are still eligible for TANF after one quarter, though most do not return once they are removed. This is consistent with the notion that families may become discouraged by onerous program participation rules and move away from attempting to claim welfare entirely.

4.4 (Un)expected Timeouts

After time limits are imposed, conditional on using up all available benefits, families know when their last day of welfare eligibility will occur. This makes welfare benefits much like Unemployment Insurance (UI) for such families. If welfare recipients are forward-looking rational agents, then they will plan for the eventual expiration of benefits by searching in the current period for employment opportunities after their benefits are exhausted, which may increase their equilibrium wage offer and potential fu-

ture wages.³⁴ Facing an expiring benefit window, a longer horizon with which to search for work pushes up equilibrium wages by increasing the reservation wage, or creating a stronger bargaining position for the welfare recipient, *ceteris paribus* (Farber and Valletta, 2015; Meyer, 1990; Schmieder et al., 2012, 2016). By contrast, families who are immediately removed from welfare cannot plan for their eventual last day on welfare. Given that these women have few savings, if they are unexpectedly removed from welfare, they may have to accept lower wages in order to reach a minimum baseline consumption in a given period. Given the persistence in wages from one period to the next, this initial offer may lead to lower lifetime consumption for such families.

We test this hypothesis in an event study framework where the *event* is reaching the 60-month federal time limit, which we call a “timeout”. The effect of the event itself is identified through variation in its timing throughout the sample. Timeouts begin happening on the day of the policy change and subsequently occur every month, though at a much lower frequency. We define an *unexpected* timeout as one that occurs on September 2011, which accounts for 50% of the total timeouts; the remainder comprise individuals who time out of the program from October 2011 onwards. We modify equation 3 by estimating

$$y_{it} = \alpha + \sum_{j=\underline{j}, j \neq -1}^{\bar{j}} \left[\beta_j b_{it}^j + \gamma_j \left(b_{it}^j \times s_{it}^j \right) \right] + \mu_i + \delta_t + \epsilon_{it} \quad (5)$$

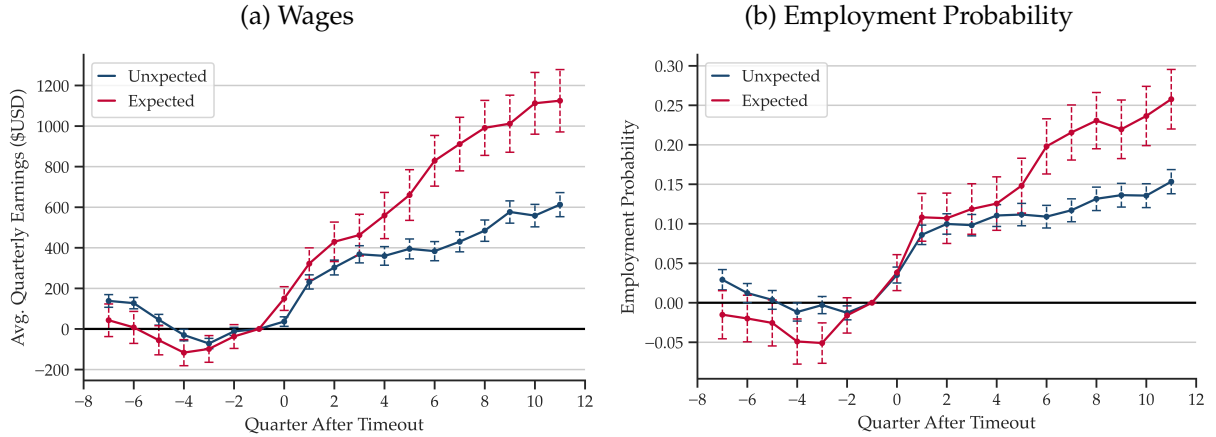
where the dummy s_{it}^j equals 1 if the event defined by b_{it}^j was a surprise to the individual. The rest of the specification follows from the original event study equation.

Given the econometric framework, someone who loses welfare with notice has a treatment effect given by β_j in a given period j after the event. Meanwhile, someone who loses welfare without notice has a treatment effect captured by $\beta_j + \gamma_j$. Thus γ_j identifies the differential treatment effect between these two events. Assuming individuals are forward looking, then – for y_{it} defined as wages or employment – we hypothesize that γ_j is negative.

Figure 8 plots the coefficients β_j and $\beta_j + \gamma_j$ for both the quarterly wage and employment probability two years before and three years after a given event. The dark blue line shows the effect of suddenly being removed, while the dark red line shows the average effect for anyone kicked off welfare due to reaching the time limit. As hypothesized, the wages and employment probabilities for the unexpected welfare leavers are strictly lower than those who could adequately plan around timing out. While the results for those who

³⁴The reasoning behind this assertion is analogous to the arguments made about the effect of increasing the UI benefit duration on the resulting wage when reentering the labor market.

Figure 8: Event Study Results For Reaching Welfare Time Limit By Circumstance



Notes: This figure shows the coefficients from estimating equation 5. The blue line is the coefficient estimates for the subsample of individuals who timed out of welfare on October 2011, while the red line is the pooled estimate of the remainder for the two outcomes, respectively.

could anticipate being removed from welfare cannot be strictly interpreted causally, they are highly suggestive and constitute an upper bound of wage and employment probabilities when exiting welfare, due to the incorporation of any forward looking behavior.

Three years after leaving welfare, these women are making \$1000 more on average per quarter relative to their welfare state. Given that the typical welfare user receives above \$400 in *monthly* benefits, the evidence suggests that on average women are not replacing their wages after timing off the program. This is even more true of the women who were not able to plan around this event. Apparent too are the low levels of labor force participation. At the peak, 3 years after timing off benefits, only a quarter of these women are working. This is evidence that many of these participants become disconnected mothers rather than working, receiving neither welfare benefits or returning to the labor market. This implies that removing them from welfare, rather than providing an incentive to replace welfare earnings with labor income, has left them in an even more precarious financial state.

5 Robustness of Main Empirical Results

Here we summarize several robustness tests and alternative specifications that we perform. Most of the analysis (and further details) are contained in appendix A.

5.1 Regression Discontinuity

As is the case in many observational studies, there are multiple suitable approaches to identify similar causal estimands. In our particular context, this is especially true, for instead of having a uniformly treated group and an untreated control group, we have a particular form of treatment assignment depending on a policy rule at the 5 year threshold. Selection on (un)observables means that a suitable treatment group could not be found from within our existing sample using, for example, individuals far from the treatment threshold who are not subject to immediate welfare removal.

An alternative approach would be to use a modified regression discontinuity design (RDD) to study the effects of the assignment rule on individuals that are as-if randomly assigned treatment at the treatment cutoff. As a robustness of our empirical design, we perform this analysis. We show that defining the average treatment effect in this way mitigates, but does not eliminate, the identification issues inherent in this study. We briefly outline this approach here, with a longer discussion in section A.1. Overall, the evidence points to the same results as those presented in section 4 and should thus be seen as a robustness check on our preferred empirical approach.

As shown in figure 2b, on September 2011 a discontinuity in the probability of being on welfare in the following period was introduced at 60 months of welfare eligibility. This fact motivates both an RDD, to test the immediate – one period – impact of losing welfare on labor market outcomes.³⁵ There are practical issues with using this approach in the current context.

The primary issue is that those to the right of the cutoff are removed from welfare, but those to the left are given a number of months remaining. This problem is compounded by the fact that our employment data is quarterly and the policy change occurs at the end of a quarter. This means that all those with 57-59 months of welfare use will either leave welfare voluntarily or become fully treated in the following period, invalidating the discontinuous treatment assignment and biasing any results towards zero. This problem becomes more severe the further into the future we wish to analyze results.

Our solution is to perform a “donut-RD” (Almond and Doyle, 2011; Barreca et al., 2011). Our aim is to estimate the immediate impact of being kicked off on employment outcomes in the following period(s). Therefore we remove those who are mechanically treated within the quarter of employment we wish to analyze to the left of the threshold

³⁵It also motivates a regression kink design (RKD). In this study, the court case – which removes the clean kink in the post period by eliminating the effects of the time limits – and sample – which contains many discrete values of the running variable and is generally underpowered – makes estimating an RKD very inconsistent. Due to high uncertainty in these results, a formal RKD is not presented in this analysis.

to account for this treatment “contamination” in this region. It is not necessary to do this to the right of the threshold given that these individuals are homogeneously treated, so this is an asymmetric donut. We find numbers that are consistent with our difference-in-differences results. The RD shows positive employment effects at the threshold of approximately 5 percentage points (15%) in the likelihood of being employed in the 4th quarter of 2011 relative to the group assigned time limits. While the affect on wages is not quite as apparent, there is much higher dispersion here, with many points well above the fitted line on the left hand side of the cutoff. The full results are provided in appendix section [A.1](#).

In the appendix section, we show robustness of these RD results along three dimensions. First we explore the impact of the bandwidth and polynomial order on the results. Next, we provide two falsification tests. We begin by demonstrating that alternative cut-off points along the same distribution do not yield statistically significant results. Finally, we repeat the same RD at each potential quarter we have data for prior to the reform. All of these tests bolster the main findings.

5.2 Dynamic Results

Here we present some informal evidence that parallel trends hold in our context by estimating our main treatment effect on losing welfare as a dynamic difference-in-differences, design. This also allows us to illustrate the effects of the court date on the treatment effects of interest. Specifically, we construct the same pre- and post-periods as defined in the primary analysis in section [4.1](#), but instead of estimating a traditional difference-in-differences regression, we estimate equation [5](#) by augmenting the original specification with flexible time dummies to capture the time variation in the overall treatment effects estimated. Doing so is illustrative; since we have placed no restrictions on any of these event dummies, coefficients that are not deemed statistically significant (relative to the quarter before treatment excluded group) indicate that the treatment and control groups do not have meaningful differences in the outcome variable prior to treatment. Furthermore, the coefficients can illuminate the heterogeneity in treatment effects by quarter following the reform.

We present the results in figure [A.3](#). The resulting estimates are highly supportive of our empirical strategy. Pre-treatment point estimates are precisely centered at zero for all variables of interest in the pre-period, a strong indication that there are no underlying time trends or reasons to doubt the parallel trends assumption. Two additional things are made clear by this analysis. First, the employment and wage effects grow over the

course of the post period, though these results are not statistically different from each other. Second, the court case mechanically increase the number of people on TANF in the post period 6 quarters after treatment. This also increases the monthly budget of these families. While it is unclear what would have happened in the absence of the court case, the evidence here indicates that there may have been larger extensive and intensive labor market responses without this attenuation of the treatment.

5.3 Alternative Control Periods

Next, we probe our identification strategy further by demonstrating that it is robust to alternative time periods chosen for the control sample. Recall that we based estimated effects on the idea that the counterfactual behavior of individuals with a counter value of D around the date of the policy change could be inferred from the behavior of families with the same counter value in the past. In our framework, identification of the difference-in-differences coefficient can control for family-level confounders; however, with this relative – event-style – type strategy for the before and after period, we cannot control for time-specific fixed effects (where there is no overlap) since there is no within-time-period variation in treatment status outside of the overlapping period to identify fixed effects. Given this fact, we may be concerned with time-dependent variation in the outcome variable that is correlated with welfare participation. This is compounded by the fact that the pre-period for the control group sample could be during the Great Recession, depending on the benchmark prior date chosen. Additionally, we may be concerned with inherent time trends in the outcome variables.

To verify that we have not cherry-picked our benchmark period chosen for the comparison, we redo our primary analysis at every hypothetical alternative control date possible. In particular, we begin with January 2009 and run the same primary analysis for both of our labor market outcomes – employment and wages – as in sections 4.1 and 4.2 and continue doing this until September 2010. These 21 counterfactual results for each counter value for the wages and employment outcomes are summarized in appendix figure A.4. Each box-plot shows the min/max and interquartile range (along with median) of the point estimates for that value depending on the counterfactual previous period chosen. We plot in red the value we show in the main text as our baseline specification. If there are outlier values, these are plotted as independent points.

Overall, the choice of previous time period does seem to affect results, but only up a point. The results we present as our main analysis are among the smallest estimated, demonstrating that we have chosen our pre-period conservatively. Importantly too, the

increasing effects closer to the 5 year horizon are true no matter the reference period, a sign that our estimation strategy has validity. In further work (not shown) we include controls for the local unemployment rate, to capture a measure of relevant time-varying heterogeneity that could bias our results. All key findings hold.

6 Conclusion

Welfare programs provide meaningful financial stability to thousands of needy families who live below the poverty line. Since the 1990s, reforms to entitlement programs, mainly through time limits and work requirements, have made benefits harder to qualify for and have nudged more and more families into the labor force. However, it is difficult to study the effects that such reforms have on the lives of affected families, because time limits are non-random in their assignment and bite years after they are assigned, and work requirements apply to all eligible recipients. This is not to mention the lack of detailed data that tracks these families over time.

To circumvent these issues, existing work relies on state-level variation in benefit generosity (or harshness in program rules) to identify the effects that such reform efforts have on the labor supply of families at the margin. While these approaches have revealed a great deal about the interplay between the generosity of benefits and labor supply, they do not form a complete picture about the lives of welfare users. Furthermore, most studies of welfare reforms are hampered by the fact that individuals are rarely – if ever – randomly removed from welfare *unexpectedly*. Rather, a small fraction of welfare recipients are timed out years after time limits are put in place, and families face the same program restrictions when they apply for benefits. Most welfare stays are short, with few reaching the time limits.

The goal of this paper has been to go beyond the normal analysis of welfare reforms and consider the wholistic effects of policies like complete removal from benefits – alongside time limits and work requirements – on the beneficiaries themselves. It did so by using administrative data to analyze a rare cleanly-identified “natural experiment” in the removal of benefits. In fact, we have found no other reforms that have created quasi-random variation in welfare receipt from one period to the next for such a large number of people as the one studied here.

We began with a simple proposition: if the goal of welfare reform is to promote self-sufficiency and employment, then do the extant tools succeed in their aim? The conclusion of this work is almost certainly that such reforms have mostly harmful consequences

on families. While we find that female heads of household do work in response to losing benefits, we find no evidence that any sizable number of them have recouped the value of their lost benefits. The second-order effects of such reforms then is likely negative. We find suggestive additional evidence that small nudges, like sanctions, that temporarily remove people from welfare rolls can keep families off these programs in significant numbers even years after the sanctions expire. Overall, affected families from all the reforms studied are almost surely poorer, on average.

Given recent work on the benefits of investing in young children and helping families move to opportunity ([Chetty et al., 2016](#); [Hendren and Sprung-Keyser, 2019](#)), alongside work that suggests policies targeted at the needy can pay for themselves ([Bastian and Jones, 2020](#)), it is unclear that making cash welfare prohibitively costly has any real policy benefits. Rather, this work, alongside other recent findings related to SNAP ([Gray et al., 2021](#)) and Medicaid ([Maclean et al., 2019](#); [Miller et al., 2018](#); [Ko et al., 2020](#)), suggests that such administrative burdens may be set at levels beyond the social welfare maximizing value.

References

- Douglas Almond and Joseph J. Doyle. After midnight: A regression discontinuity design in length of postpartum hospital stays. *American Economic Journal: Economic Policy*, 3(3):1–34, aug 2011. ISSN 19457731. doi: 10.1257/pol.3.3.1. URL <http://www.aeaweb.org/articles.php?doi=10.1257/pol.3.3.1><http://www.aeaweb.org/articles.php?doi=10.1257/pol.3.3.1>.
- Alan I. Barreca, Melanie Guldi, Jason M. Lindo, and Glen R. Waddell. Saving babies? Revisiting the effect of very low birth weight classification. *Quarterly Journal of Economics*, 126(4):2117–2123, nov 2011. ISSN 00335533. doi: 10.1093/qje/qjr042. URL <https://academic.oup.com/qje/article/126/4/2117/1924019>.
- Jacob Bastian and Maggie R Jones. Do EITC Expansions Pay for Themselves? Effects on Tax Revenue and Public Assistance Spending. *Working Paper*, 2020. URL <https://drive.google.com/file/d/1GbBeeQzfGH9fF9Y1u5rS55Sn3eStBWE7/view>.
- Neil G Bennett, Hsien-Hen Lu, and Younghwan Song. Welfare Reform and Changes in the Economic Well-being of Children. *Population Research and Policy Review*, 23(5-6):671–699, 2004. URL <https://link.springer.com/article/10.1007/s11113-004-2930-3>.
- Marianne P Bitler, Jonah B Gelbach, and Hilary W Hoynes. What Mean Impacts Miss: Distributional Effects of Welfare Reform Experiments. *American Economic Review*, 96(4):988–1012, sep 2006. URL <http://www.aeaweb.org/articles?id=10.1257/aer.96.4.988>.
- Rebecca M Blank. Evaluating Welfare Reform In The United States. *Journal of Economic Literature*, 2002. ISSN 00220515. doi: 10.1257/.40.4.1105.
- Rebecca M Blank. Improving the Safety Net for Single Mothers Who Face Serious Barriers to Work. *The Future of Children*, 17(2):183–197, dec 2007. ISSN 10548289, 15501558. URL <http://www.jstor.org/stable/4495066>.
- Rebecca M Blank and Brian K Kovak. The Growing Problem Of Disconnected Single Mothers. In *Making the Work-Based Safety Net Work Better: Forward-Looking Policies to Help Low-Income Families*. 2009. ISBN 9780871544667.
- Sebastian Calonico, Matias D Cattaneo, Max H Farrell, and Rocío Titiunik. rdrobust: Software for regression-discontinuity designs. Technical Report 2, 2017.

- Frances Carley. The Family Independence Program (FIP): 48-Month and 60-Month Time Limits. Technical report, Michigan Senate Fiscal Agency, 2011.
- Matias D. Cattaneo, Nicolás Idrobo, and Rocío Titiunik. *A Practical Introduction to Regression Discontinuity Designs*. Cambridge University Press, nov 2019. ISBN 9781108684606. doi: 10.1017/9781108684606. URL <https://www.cambridge.org/core/product/identifier/9781108684606/type/element>.
- Marc K Chan. A Dynamic Model of Welfare Reform. *Econometrica*, 81(3):941–1001, 2013. ISSN 0012-9682. URL <https://onlinelibrary.wiley.com/doi/abs/10.3982/ECTA9075>.
- Raj Chetty, Nathaniel Hendren, and Lawrence F Katz. The Effects of Exposure to Better Neighborhoods on Children: New Evidence from the Moving to Opportunity Experiment. *American Economic Review*, 106(4):855–902, 2016. doi: 10.1257/aer.20150572. URL <https://www.aeaweb.org/articles?id=10.1257/aer.20150572>.
- Clément de Chaisemartin and Xavier D’Haultfoeuille. Two-Way Fixed Effects Estimators with Heterogeneous Treatment Effects. *American Economic Review*, 110(9):2964–2996, sep 2020. doi: 10.1257/aer.20181169. URL <https://www.aeaweb.org/articles?id=10.1257/aer.20181169>.
- Manasi Deshpande. Does Welfare Inhibit Success? The Long-Term Effects of Removing Low-Income Youth from the Disability Rolls. *American Economic Review*, 106(11):3300–3330, nov 2016a. URL <http://www.aeaweb.org/articles?id=10.1257/aer.20151129>.
- Manasi Deshpande. The Effect of Disability Payments on Household Earnings and Income: Evidence from the SSI Children’s Program. *The Review of Economics and Statistics*, 98(4):638–654, 2016b. URL https://doi.org/10.1162/REST_a_00609.
- Nada Eissa and Hilary Williamson Hoynes. Taxes and the Labor Market Participation of Married Couples: The Earned Income Tax Credit. *Journal of Public Economics*, 88(9):1931–1958, 2004. ISSN 0047-2727. URL <http://www.sciencedirect.com/science/article/pii/S0047272703001440>.
- Nada Eissa and Austin Nichols. Tax-Transfer Policy and Labor-Market Outcomes. *American Economic Review*, 2005. ISSN 00028282. doi: 10.1257/000282805774670527.
- Henry S. Farber and Robert G. Valletta. Do Extended Unemployment Benefits Lengthen Unemployment Spells? Evidence from Recent Cycles in the U.S. Labor Market. *Journal of Human Resources*, 50(4):873–909, oct 2015. ISSN 0022166X. doi: 10.3368/jhr.

50.4.873. URL <http://jhr.uwpress.org/content/50/4/873><http://jhr.uwpress.org/content/50/4/873.abstract>.

Sergio Firpo, Nicole M Fortin, and Thomas Lemieux. Unconditional Quantile Regressions. *Econometrica*, 77(3):953–973, jun 2009. ISSN 00129682, 14680262. URL <http://www.jstor.org/stable/40263848>.

Ron French. Welfare Reform Leaves Families Without A Net, And Off The Radar. *Bridge Magazine*, feb 2012. URL <https://www.bridgemi.com/children-families/welfare-reform-leaves-families-without-net-and-radar>.

Phil Galewitz. Trump officials say no to lifetime limits on Medicaid. *CNN Money*, 2018. URL <https://money.cnn.com/2018/05/08/news/economy/lifetime-limits-medicaid/index.html>.

John Gardner. Two-Stage Differences in Differences. *Working Paper*, (April), 2021.

Rachel Garfield, Robin Rudowitz, Kendal Orgera, and Anthony Damico. Understanding the Intersection of Medicaid and Work: What Does the Data Say? *Kaiser Family Foundation Issue Brief*, 2019. URL <https://www.kff.org/medicaid/issue-brief/understanding-the-intersection-of-medicaid-and-work-what-does-the-data-say/>.

Craig Garthwaite, Tal Gross, and Matthew J Notowidigdo. Public Health Insurance, Labor Supply, and Employment Lock. *The Quarterly Journal of Economics*, 129(2):653–696, 2014. ISSN 0033-5533. URL <https://doi.org/10.1093/qje/qju005>.

Olivia Golden and Amelia Hawkins. TANF Child-Only Cases. Technical report, Urban Institute: Office of Planning, Research and Evaluation, 2011.

Andrew Goodman-Bacon. Difference-in-differences with variation in treatment timing. *Journal of Econometrics*, 2021. ISSN 0304-4076. doi: <https://doi.org/10.1016/j.jeconom.2021.03.014>. URL <https://www.sciencedirect.com/science/article/pii/S0304407621001445>.

Colin Gray, Adam Leive, Elena Prager, Kelsey Pukelis, and Mary Zaki. Employed In A Snap? The Impact Of Work Requirements On Program Participation And Labor Supply. *Work in Progress*, 2019. URL <https://economics.mit.edu/files/17896>.

Colin Gray, Adam Leive, Elena Prager, Kelsey B Pukelis, and Mary Zaki. Employed in a SNAP? The Impact of Work Requirements on Program Participation and Labor Supply. *National Bureau of Economic Research Working Paper Series*, No. 28877, 2021.

doi: 10.3386/w28877. URL <http://www.nber.org/papers/w28877><http://www.nber.org/papers/w28877.pdf>.

Jeffrey Grogger. 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, 2003. URL <https://doi.org/10.1162/003465303765299891>.

Jeffrey Grogger. Time limits and welfare use. *Journal of Human Resources*, 2004. ISSN 0022166X. doi: 10.2307/3559020.

Jeffrey Grogger and Lynn A Karoly. Welfare Reform: Effects Of A Decade Of Change. *Choice Reviews Online*, 2006. ISSN 0009-4978. doi: 10.5860/choice.43-4782.

Jeffrey Grogger and Charles Michalopoulos. Welfare Dynamics Under Time Limits. *Journal of Political Economy*, 111(3):530–554, 2003. URL <https://doi.org/10.1086/374181>.

Jeffrey Grogger, Steven J Haider, and Jacob Klerman. Why Did the Welfare Rolls Fall During the 1990’s? The Importance of Entry. In *American Economic Review*, 2003. doi: 10.1257/000282803321947218.

Nathaniel Hendren and Benjamin D Sprung-Keyser. A Unified Welfare Analysis Of Government Policies. *NBER Working Paper*, 2019. ISSN 0898-2937. URL <https://www.nber.org/papers/w26144>.

Andrea Hetling, Karen Baehler, and Rafay Kazmi. Federal Welfare Time-Limit Extensions and Exemptions: Why Does Utilization Vary across States and over Time? *Social Service Review*, 94(3):567–606, sep 2020. ISSN 0037-7961. doi: 10.1086/710556. URL <https://doi.org/10.1086/710556>.

Hilary Williamson Hoynes and Diane Whitmore Schanzenbach. Work Incentives and the Food Stamp Program. *Journal of Public Economics*, 96(1):151–162, 2012. ISSN 0047-2727. URL <http://www.sciencedirect.com/science/article/pii/S0047272711001472>.

Michael P Keane and Kenneth I Wolpin. The Role Of Labor And Marriage Markets, Preference Heterogeneity, And The Welfare System In The Life Cycle Decisions Of Black, Hispanic, And White Women. *International Economic Review*, 51(3):851–892, 2010. ISSN 0020-6598. URL <https://www.jstor.org/stable/40784808>.

Henrik Kleven. The EITC and the Extensive Margin: A Reappraisal. *NBER Working Paper*, (26405), oct 2019. URL <http://www.nber.org/papers/w26405>.

- Patrick Kline and Melissa Tartari. Bounding the Labor Supply Responses to a Randomized Welfare Experiment: A Revealed Preference Approach. *American Economic Review*, 106(4):972–1014, 2016. ISSN 0002-8282. URL <https://www.aeaweb.org/articles?id=10.1257/aer.20130824>.
- Hansoo Ko, Renata E Howland, and Sherry A Glied. The Effects of Income on Children’s Health: Evidence from Supplemental Security Income Eligibility under New York State Medicaid. *NBER Working Paper*, (26639), jan 2020. URL <http://www.nber.org/papers/w26639>.
- Emily Lawler. Welfare Time Limits Save Michigan Millions, But Cost 32,090 Families. *MLive*, apr 2016. URL https://www.mlive.com/news/2016/04/saving_dollars_not_people_chan.html.
- Hamish Low, Costas Meghir, Luigi Pistaferri, and Alessandra Voena. Marriage, Labor Supply and the Dynamics of the Social Safety Net. *NBER Working Paper*, 2018. URL <https://www.nber.org/papers/w24356>.
- Elizabeth Lower-Basch. TANF Policy Brief: Guide to Use of Funds. Technical report, Center for Law and Social Policy, Washington, DC, 2011.
- Johanna Catherine Maclean, Sebastian Tello-Trillo, and Douglas Webber. Losing Insurance and Behavioral Health Hospitalizations: Evidence from a Large-scale Medicaid Disenrollment. *NBER Working Paper*, (25936), jun 2019. URL <http://www.nber.org/papers/w25936>.
- E Meara and R Frank. Welfare Reform, Work Requirements, And Employment Barriers, 2006.
- Bruce D Meyer. Unemployment Insurance and Unemployment Spells. *Econometrica*, 58(4):757–782, 1990. URL <https://ideas.repec.org/a/ecm/emetrp/v58y1990i4p757-82.html>.
- Bruce D Meyer and Dan T Rosenbaum. Welfare, the Earned Income Tax Credit, and the Labor Supply of Single Mothers*. *The Quarterly Journal of Economics*, 116(3):1063–1114, aug 2001. ISSN 0033-5533. doi: 10.1162/00335530152466313. URL <https://doi.org/10.1162/00335530152466313>.
- Michigan DHHS. Information Packet - August 2011. 2011. URL https://www.michigan.gov/documents/dhs/DHS-PUB-0790-August2011_362418_7.pdf.

- Michigan DHHS. Michigan's Work Participation Rate Hits Record High, 2014. URL https://www.michigan.gov/mdhhs/0,5885,7-339-73970%7B%5C_%7D71692%7B%5C_%7D7752-338116--,00.html.
- Sarah Miller, Luojia Hu, Robert Kaestner, Bhashkar Mazumder, and Ashley Wong. The ACA Medicaid Expansion in Michigan and Financial Health. *NBER Working Paper*, (25053), sep 2018. URL <http://www.nber.org/papers/w25053>.
- Robert A Moffitt. The Temporary Assistance for Needy Families Program. In *Means-tested Transfer Programs in the United States*, pages 291–364. University of Chicago Press, 2003. URL <https://www.nber.org/chapters/c10258.pdf>.
- Laura Reiley. Trump administration tightens work requirements for SNAP, which could cut hundreds of thousands from food stamps. *Washington Post*, 2019.
- Sebastian Schmidheiny and Kurt Sieglöcher. On Event Studies and Distributed-Lags in Two-Way Fixed Effects Models: Identification, Equivalence, and Generalization. *ZEW Discussion Papers*, No. 20-017, 2020.
- Johannes F. Schmieder, Till Von Wachter, and Stefan Bender. The Long-Term Effects of UI Extensions on Employment. In *American Economic Review*, volume 102, pages 514–519, 2012. doi: 10.1257/aer.102.3.514.
- Johannes F. Schmieder, Till von Wachter, and Stefan Bender. The Effect of Unemployment Benefits and Nonemployment Durations on Wages. *American Economic Review*, 106(3):739–777, 2016. doi: 10.1257/aer.20141566. URL <http://dx.doi.org/10.1257/aer.20141566>.
- Liz Schott, LaDonna Pavetti, and Ife Floyd. How States Use Federal and State Funds Under the TANF Block Grant. Technical report, Center on Budget and Policy Priorities, 2015. URL <https://www.cbpp.org/sites/default/files/atoms/files/4-8-15tanf.pdf>.
- Christopher A Swann. Welfare Reform When Recipients Are Forward-Looking. *Journal of Human Resources*, XL(1):31–56, jan 2005. doi: 10.3368/jhr.XL.1.31. URL <http://jhr.uwpress.org/content/XL/1/31.abstract>.
- Lesley J Turner, Sheldon Danziger, and Kristin S Seefeldt. Failing the Transition from Welfare to Work: Women Chronically Disconnected from Employment and Cash Welfare. *Social Science Quarterly*, 2006. ISSN 00384941. doi: 10.1111/j.1540-6237.2006.00378.x.

U.S. DHHS. Characteristics and Financial Circumstances of TANF Recipients, Fiscal Year 2011. 2013. URL <https://www.acf.hhs.gov/ofa/resource/characteristics-financial-circumstances-appendix-fy2011>.

James P Ziliak. *Welfare Reform and Its Long-Term Consequences for America's Poor*. 2009. ISBN 9780511605383. doi: 10.1017/CBO9780511605383.

James P Ziliak. Temporary Assistance For Needy Families. In Robert A Moffitt, editor, *Economics Of Means-tested Transfer Programs In The United States, Volume 1*, pages 303–393. University of Chicago Press, 2015. URL <https://www.nber.org/chapters/c13483>.

A Appendix

A.1 Regression Discontinuity Results

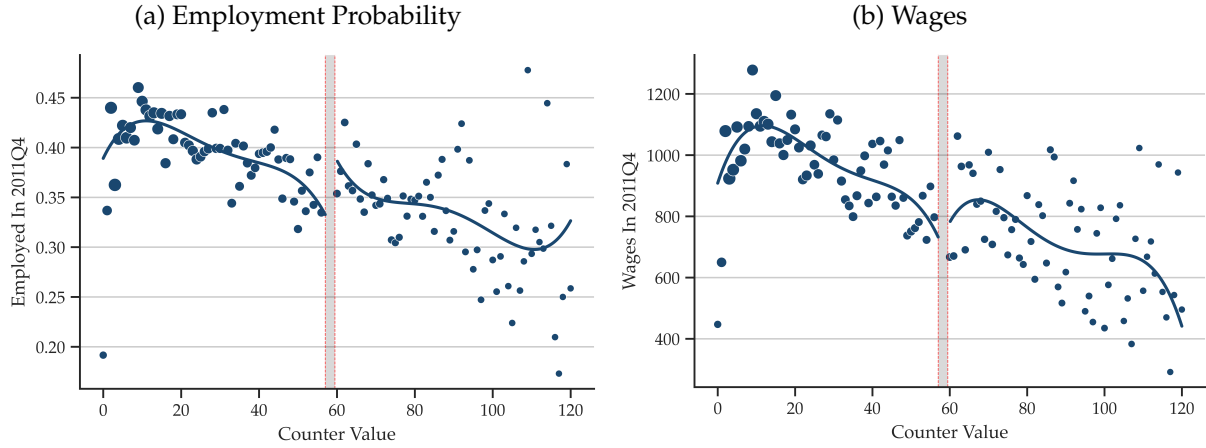
A.1.1 Main Results

We begin with the graphical results which should be interpreted with caution due to the discrete running variable, polynomial order chosen, and bandwidth selected. For simplicity here and throughout, we employ the defaults of standard regression discontinuity packages: a fourth order polynomial on either side of the cutoff and binning the running variable using mimicking variance evenly-spaced bins (Calonico et al., 2017). We use a parabolic (Epanechnikov) kernel. We add additional clarity by adjusting the size of each point in the scatterplot to represent the amount of mass at that point: the larger the dot the more individuals are represented at that point.

The results on employment for both wages and the probability of having positive formal earnings are shown in figure A.1. Both show what appears to be a positive effect at the threshold consistent with the results in section 4. In particular, there appears to be a positive employment effect at the threshold of approximately 5 percentage points (15%) in the likelihood of being employed in the 4th quarter of 2011 relative to the group assigned time limits. While the affect on wages is not quite as apparent, there is much higher dispersion here, with many points well above the fitted line on the left hand side of the cutoff. Moreover, the treatment at the cutoff is not perfectly assigned; when divided by the “first stage” (figure 2b), these estimates will grow in magnitude when attaining the average treated effect.

Rather than rely on visual results – which by their nature are highly sensitive to the functional form of the polynomial (both order and support) in addition to the binning of the running variable – and to estimate the jump at the cutoff empirically – incorporating the first stage effects of removal from welfare and covariates to increase precision and calculate the average treatment effect of the policy – we run local linear regressions. In table A.1 we show the results of local linear regressions of the discontinuities where bandwidths have been selected using an MSE-optimal procedure adjusting for mass points in the running variable (Calonico et al., 2017). We use a parabolic (Epanechnikov) kernel, though results are nearly identical regardless of kernels chosen (uniform or triangular). Cluster robust nearest-neighbor standard errors are shown in parentheses, with the neighbors set to 3. We add covariates to increase precision. While they do not change the point estimates themselves, they add precision to the standard errors, particularly those that control for previous employment status. The full list of covariates include dummies for

Figure A.1: Regression Discontinuity Plots For Oct-Dec 2011



Notes: This figure shows the RDD of interest as described in the main text for employment and wages in the quarter after treatment. We use we employ the defaults of standard regression discontinuity packages: a fourth order polynomial on either side of the cutoff and binning the running variable using mimicking variance evenly-spaced bins (Calonico et al., 2017). The shaded region is excluded and represents the individuals who would become treated in the quarter after the event.

race, ethnicity, whether the individual lives in Detroit, whether their education is above or below high school graduation, and whether they are employed prior to the reform, along with continuous controls for wages prior to the reform, number of children, and age. We explore the impact of the bandwidth choice and changing to a parametric estimator in the following subsection.

The first column presents the first stage estimate of welfare use in the quarter following the policy change at the cutoff. It is a slightly smaller estimate of the drop in probability from the estimates of the full drop shown in figure 2b, because we have recomputed TANF participation at the quarterly level to indicate whether you are observed on TANF at the end of 2011 to be consistent with the outcome variables of interest. Each subsequent column shows the effect on an outcome variable, showing both the reduced form estimate and the Wald IV estimate using TANF participation as a first stage. Consistent with the RD plot shown in figure A.1, the reduced form estimate in fourth quarter employment is 4 percentage points. When rescaled by rescaled by the first stage, this implies a 7 percentage point increase in formal labor supply due to unexpected removal from welfare. However, there is no statistically significant increase in wages at the cut-off in this baseline specification, despite positive point estimates of roughly \$90. Finally, the average monthly budget drops by \$294 in the reduced form and roughly \$450 when rescaled by the first stage. This is slightly smaller than the maximum monthly benefit for a family of 3 – the modal family – of \$492. While this is, in part, mechanical due to the drop in grant, it is notable that despite increases in employment, the increase in wages,

Table A.1: Regression Discontinuity Main Results

	On Welfare	Employed		Wages		Budget	
		RF	IV	RF	IV	RF	IV
Estimate	−0.65** (0.02)	0.05* (0.02)	0.07* (0.03)	89.59 (53.59)	119.58 (80.65)	−293.73** (22.71)	−452.72** (30.14)
Bandwidth	17.09	21.52	21.77	23.31	22.54	22.02	22.77
N (Left)	3785	5120	5120	5905	5507	5507	5507
N (Right)	3179	3737	3737	4002	3876	3876	3876
N	35304	35304	35304	35304	35304	35304	35304

Notes: This table shows results from local linear regressions from the RD Robust package of [Calonico et al. \(2017\)](#). The bandwidth is selected using MSE-optimal bandwidth selection. We use a parabolic (Epanechnikov) kernel. The first column shows the sharp result of welfare participation at the cutoff. The remainder show reduced form (RF) estimates at the cutoff and fuzzy (IV) estimates using on welfare as an instrument. Estimates include robust bias-corrected standard errors clustered by nearest neighbor (3 clusters, the default). Covariates are included to more precisely estimate standard errors and include dummies for educational attainment (less than high school, more than high school) and whether their county of residence is Detroit, along with continuous controls for employment (probability and wages) in the third quarter of 2011, and age and number of children of the participant. ** Indicates p -value is less than .01, * Indicates p -value is less than .05, + Indicates p -value is less than .1.

employment, and food stamp benefits does not make up for the loss in welfare benefits for these families.

A.1.2 Robustness

Since our running variable is discrete and has a small gap due to the donut estimation, it is likely to be particularly sensitive to the bandwidth selected. Furthermore with mass points and missing mass around the cutoff due to the donut, nonparametric estimation may not be applicable.³⁶ To investigate all these concerns, we present in table A.2 the reduced form results for each of the three main outcome variables at bandwidths in increments of 10 along the distribution of the running variable. Each cell is the point estimate on the jump estimated at the cutoff of 60 months using either a linear or polynomial regression within a given bandwidth of the cutoff. Specifically we estimate, at various bandwidths and polynomial orders, the following equation

$$y_i = \alpha + \beta_0 T_i + \beta_1 r_i + \beta_2 (r_i \times T_i) + \beta_3 r_i^2 + \beta_4 (r_i^2 \times T_i) + X_i + \epsilon_i$$

where r_i is the running variable (re-centered at 0) for individual i , T_i indicates whether that individual is to the right of the cutoff, y_i is a particular outcome, and X_i are covariates

³⁶As noted in [Cattaneo et al. \(2019\)](#), when the running variable is discrete with few mass points, local polynomial methods essentially reduce each mass point to a single observation.

Table A.2: Sensitivity of RD Results to Bandwidth and Polynomial Order

	Employed		Wages		Budget	
BW	Linear	Quadratic	Linear	Quadratic	Linear	Quadratic
10	0.02 (0.03)	0.07 (0.11)	−26.66 (107.34)	198.83 (381.67)	−316.72** (43.03)	−210.45 (152.84)
20	0.05** (0.02)	0.03 (0.04)	114.45* (55.28)	12.31 (117.75)	−288.7** (22.36)	−313.71** (46.85)
30	0.04** (0.01)	0.05+ (0.02)	124.43** (42.32)	60.87 (74.4)	−284.97** (17.16)	−304.4** (30.17)
40	0.02* (0.01)	0.05** (0.02)	101.84** (35.95)	112.83+ (58.4)	−304.55** (14.52)	−274.53** (23.81)
50	0.03** (0.01)	0.03+ (0.02)	113.56** (31.96)	98.16+ (50.14)	−309.25** (12.84)	−283.76** (20.31)
60	0.03** (0.01)	0.03* (0.01)	141.14** (29.21)	74.69+ (45.12)	−301.69** (11.75)	−306.2** (18.25)

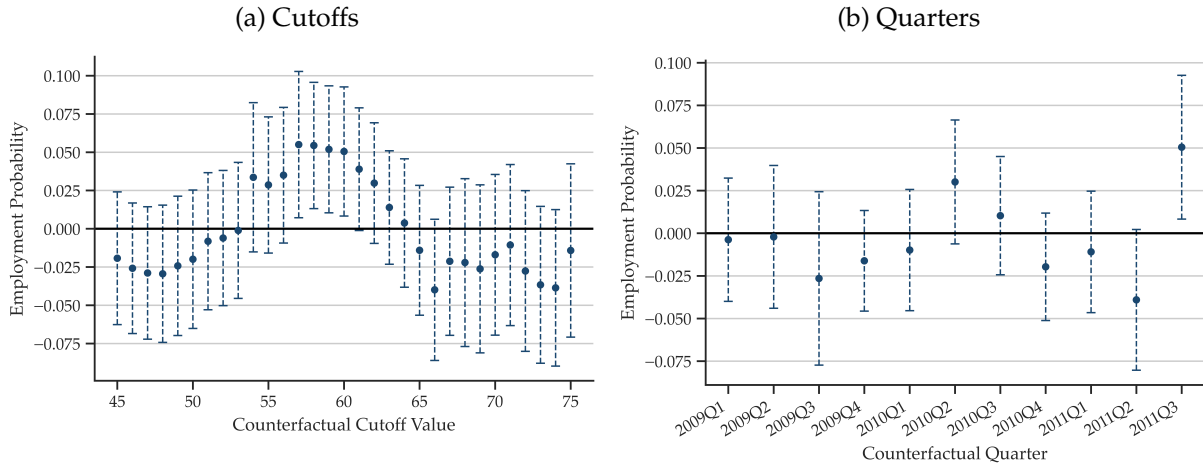
Notes: This table shows first- and second-order polynomial regression discontinuity estimates at various bandwidths. Each cell is the point estimate of the discontinuity at a particular bandwidth (in increments of 10) along with heteroskedasticity-robust standard errors in parentheses. ** Indicates p -value is less than .01, * Indicates p -value is less than .05, + Indicates p -value is less than .1.

added for precision. The widest uses the entire distribution from 0 to 120 months on TANF, while the narrowest only uses the data from 50 to 70 months. When the linear model is estimated, we omit estimation of the β_3 and β_4 terms.

In general the result for the employment probability and the total budget of the welfare participants at the threshold is highly robust to variation in polynomial order and bandwidth. Estimates of the discontinuity for wages vary based on the bandwidth and the functional form chosen. However, the magnitudes roughly stabilize as the bandwidths increase to be approximately \$100. Results from the parametric estimation are, in general, in line with the nonparametric empirical methods. They imply a reduced form estimate of roughly 4 percentage points in formal employment due to the policy change, an increase in quarterly wages of \$100 – imprecisely estimated – and a total drop in monthly budget of \$300.

Next, we provide two falsification tests of the main empirical estimates for employment shown in table A.1. We focus on the result for employment since the estimates for wages were only statistically significant at large bandwidths. The goal is to test the robustness of the result we are presenting as a valid estimate of the policy change. The first takes as given the assumption that we could have randomly chosen a cutoff anywhere along the distribution shown in figure A.1 and obtained positive estimates with

Figure A.2: Counterfactual Regression Discontinuity Estimates By Type



Notes: This figure shows the results of two counterfactual RDD exercises. The first holds constant the sample and estimates the treatment effect of interest non-parametrically but for each potential hypothetical other cutoff value. The second does the same analysis at the same cutoff value but for each potential quarter of analysis prior to the estimates shown in the main text (at the end).

our donut estimation. This is a reasonable counter argument to the results presented thus far since without fitted lines and shading to highlight the discontinuity, the jump in employment probability is not readily apparent. To address these concerns, we run counterfactual donut regression discontinuity estimates for each potential cutoff value in a wide bandwidth of the main cutoff. We employ the exact same estimation strategy as those shown in table A.1. We summarize the results in a graph where each point and accompanying 95% confidence interval give the fitted value from the nonparametric estimation on the discontinuity of the donut RD with the x-axis value as the chosen cutoff. Figure A.2a presents the results. As shown, only the counterfactual cutoffs near the true value of 60 are statistically significant. It is reasonable that several values below the true cutoff are also statistically significant, since those individuals with counters just below 60 are also kicked off in the following quarter – precisely the rationale for the donut RD.

It is also possible that the 5 year mark changes the probability, but that this “shock” was in fact nonrandom. This would be the case if Michigan had been enforcing its five year time limit in indirect ways, by increasing enforcement of work requirements and other program rules for individuals above the five year mark. If this were the case, then the effects on employment seen at the threshold would be not due to randomly being removed from welfare, but rather to behavioral effects that accompany expected loss of welfare. To test this hypothesis, we repeat the same exercise at every quarter prior to the true reform date and plot the results in figure A.2b.³⁷ As expected, each quarter prior

³⁷We do not do the same for those after this date, since we know for certain that there is true cutoff at

to the true quarter of the policy does not have statistically significant results; moreover, most point estimates are fairly close to 0.

A.2 Robustness of Labor Market Results

In this section, we provide the main robustness checks on our primary empirical design. For missing descriptions, please see the main text section 5.

A.2.1 Dynamic Difference-in-Differences

In this subsection, we run the main difference-in-differences specification with event time dummies at a monthly- or quarterly-frequency to see if there are pre-trends in treatment detectable in the year before treatment. Specifically, we estimate

$$y_{it} = \alpha + \sum_{j=\underline{j}, j \neq -1}^{\bar{j}} \beta_j b_{it}^j + \mu_i + \delta_{\tilde{t}} + \epsilon_{it}$$

with (relative) event time dummies b_{it}^j , unit fixed effects, μ_i , and relative event time dummies $\delta_{\tilde{t}}$. We exclude the period before treatment. In the case of the quarterly data, event period zero includes the months of July-September (the third quarter of the calendar year). We exclude event time -1, the month, or quarter, before treatment. Results are shown in figure A.3.

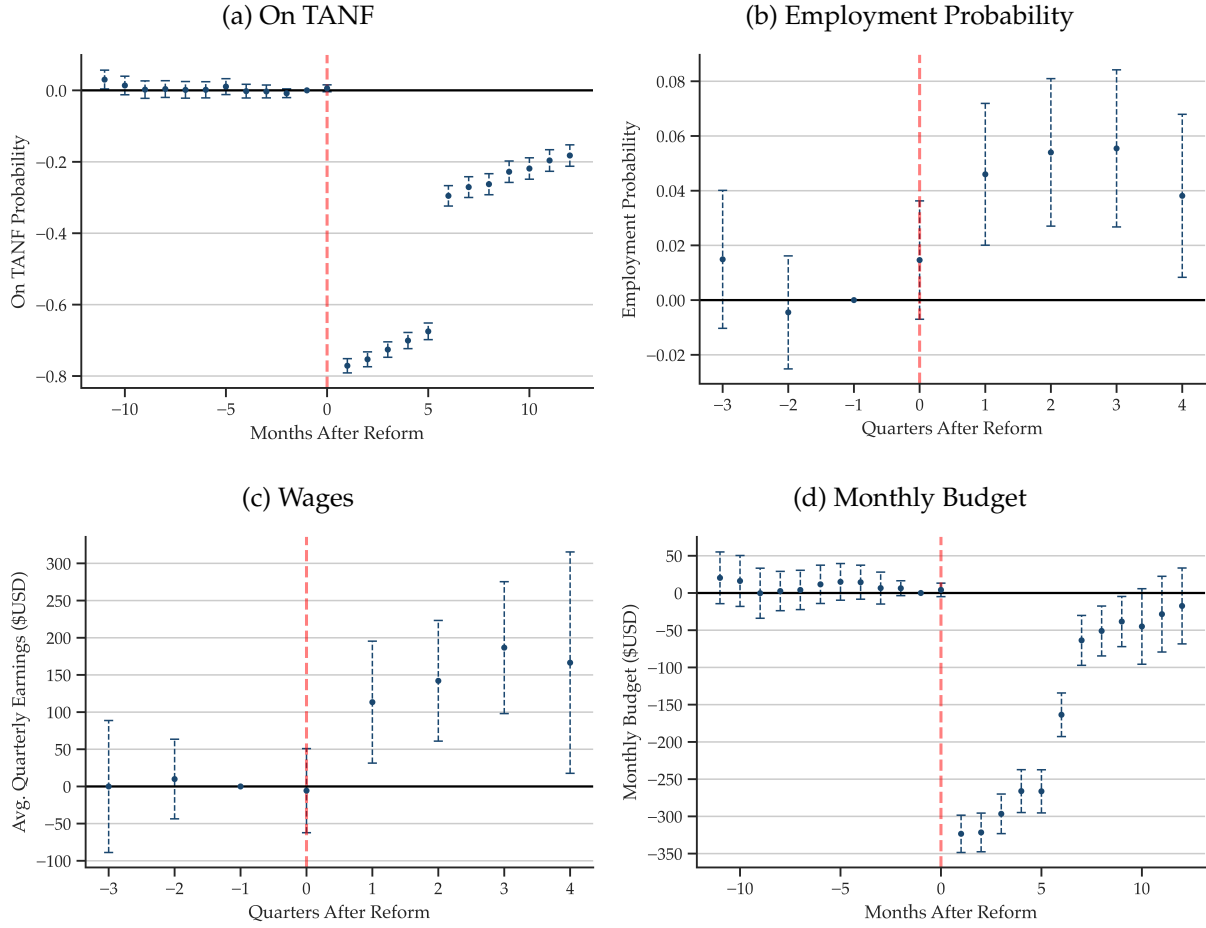
A.2.2 Counterfactual Dates

Figure A.4 presents the results of doing our relative-time difference-in-differences approach on each counterfactual prior reference date. Each box and whisker plot summarizes 21 alternative choices of past dates as a reference point for the main analysis at the respective counter value (years on welfare). The whiskers on the plot give the min and max point estimates, while the box gives the interquartile range, along with the median estimate.

The values we present as our main result – the closest possible date such that we have one full year of post period tend to be the most conservative ones estimated. However, standard errors are not shown and most estimated values cannot be said to be statistically significantly different. The reason this value is the most conservative may be because the

this point, and there are no longer individuals on welfare above the 60 month counter value to identify the discontinuity at the cutoff value.

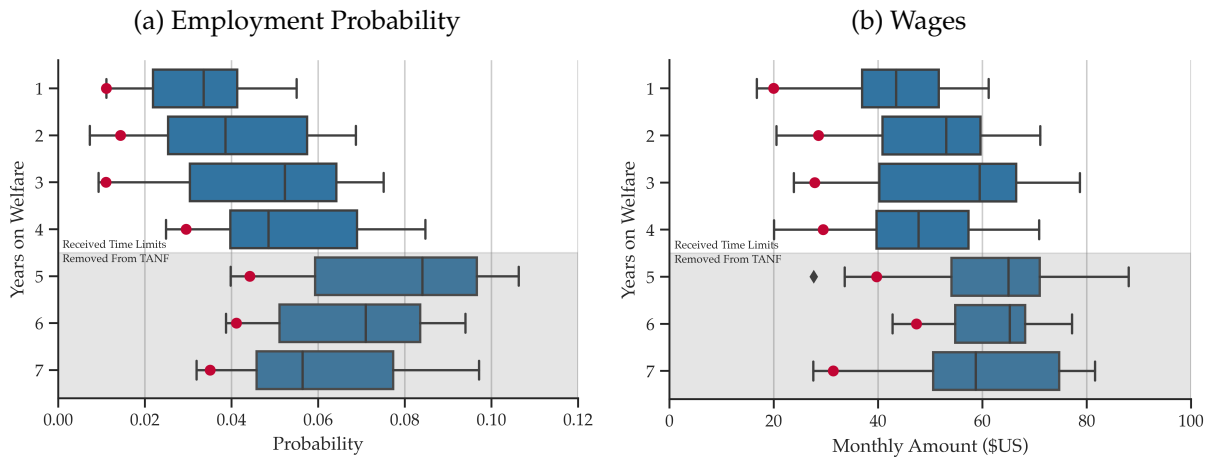
Figure A.3: Dynamic Specification



Notes: This figure shows the estimated event study dummy coefficients from estimating the main equation as a dynamic difference-in-differences. We exclude the event time of -1 so that all coefficients can be interpreted relative to this period.

post period of the control group may include some anticipatory effects of the policy – given that the policy was announced in September, and therefore, one month of the post-period could have been spent in employment in anticipation of removal. It also may be because it is less likely to have time trends or time-dependent confounders. Regardless, the estimates largely stand to confirm each other, and while the differences may seem large in relative scale when presented in this way, these are narrow differences between estimates in over magnitudes (4% vs. 8% employment gains, and \$40 vs. \$60 a month in earnings for the main treatment estimate, when comparing to the median, respectively).

Figure A.4: Robustness To Counterfactual Past Dates



Notes: Here we show the same results meant to test the robustness of those shown in figure 6. Using the same definitions of the treated group, we redo the analysis using each alternative control group available to us by month in the past (for each y -axis counter value). The red dots show the main point estimate we showed in the text. The box-and-whisker plots show the min, max, interquartile range, median, and outliers of the point estimates found in this analysis.

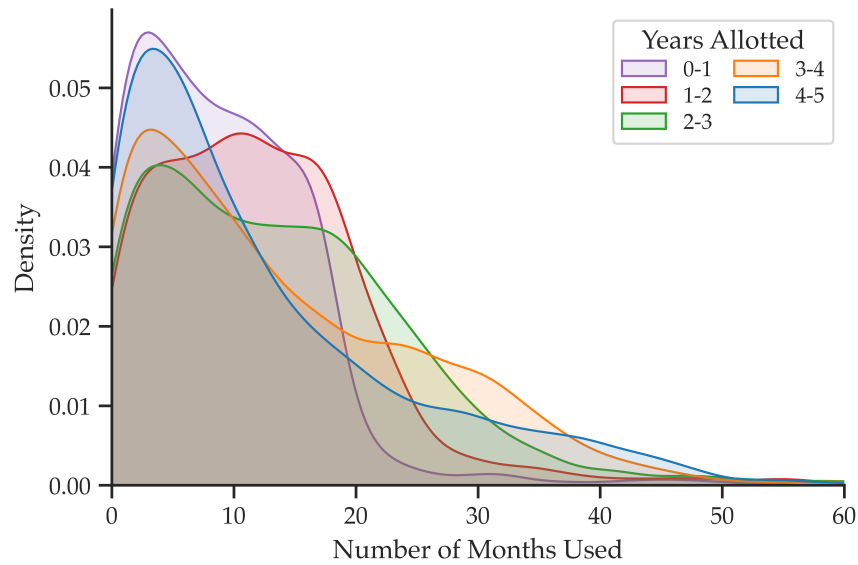
A.3 Time Limits

A.3.1 Variation in Assigned Limits and Usage

Figure A.5 shows the distribution – using a standard kernel density estimator – of months used by families assigned 0-1, 1,2, 2-3, 3-4, and 4-5 years remaining on welfare. As you can see the distributions overlap significantly. This figure includes the period during which the court injunction 6 months after the initial policy that put families who would have been affected back onto benefits. That is why the distribution of use for those assigned 0-1 years on welfare extends outward past 12 months. This drops off precipitously at around 20, which makes sense since the court case added about 6-8 months of eligibility for these families. As for those families unaffected by the court case, these distributions have a fat left tail; usage at all levels of assignment remains mostly below 10 months of welfare, even for those given 60 months of usage.

The previous figure is absolute terms, but how do the distributions compare in the *percentage* of their allotted benefits utilized? We see that families assigned 3 years use about the same number of benefits as those assigned 4 years – 9 months – but the latter only use about 20% of their benefit allotment, while the former use 25%. To compare these families more directly, figure A.6 shows the same data, but in terms of the fraction of the total assigned benefits utilized by each assignment-type. The figure shows a similar pattern as that of figure 5 while also telling us far more about the typical family on welfare. Beginning with the dark purple and red distributions, we can see that both have fat right

Figure A.5: The Distribution of Benefits Used By Years Allotted

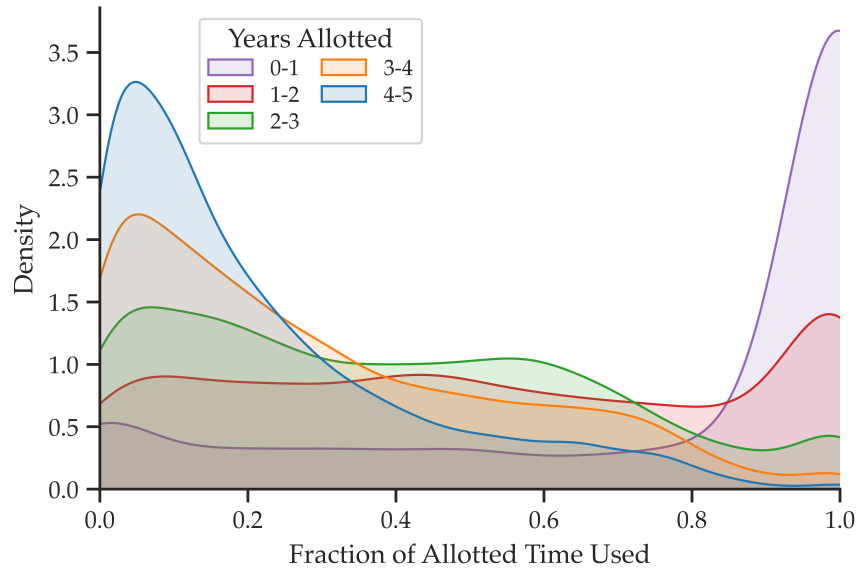


Notes: This figure shows the distribution of welfare utilization from October 2011 to December 2019 by time limit assignment separately for 5 groupings of families on welfare (0-1,1-2,...,4-5 years of welfare assigned as a result of the policy in 2011). Time limits are assigned based on prior use, so by definition those with 3 years assigned have already utilized 2 years of benefits.

tails. Families assigned fewer than 2 years have a disproportionate number of families at, or nearing, benefit exhaustion levels. However, it is also surprising that the mean showed in figure 5 hides the fact that a relatively uniform number of families use between 0 and 19 (~20%) months of benefits. For those assigned 3 or more years, the distributions gain increasingly fatter left tails. To the right, the distributions of those with more than 3 years of benefits look nearly identical.

As noted, significant variation in benefits exists in the amount of months lost due to the age of youngest child of the participant by numbers of months used. To see this variation more clearly, using the same color schemes as in A.6 (though with different labels since this is years on vs. years assigned) figure A.7 shows the variation in months lost as a function of time on welfare. Shown are violin-plots that displays the full distributions in comparable form. The shape illustrates the relative mass of observations, while the dotted lines in each subfigure show the median and interquartile ranges. These distributional statistics are close to equal between the families, and the distributions are similar, even though women with more months on welfare mechanically lose more eligibility, assuming they have children of approximately the same youngest age. Regardless, the specification of our linear model implicitly uses variation between individuals with the exact same number of months used, to isolate the policy-relevant variation of eligibility lost between families with similar prior use. Most women lost between 100 and

Figure A.6: The Distribution of Benefits Used By Years Allotted



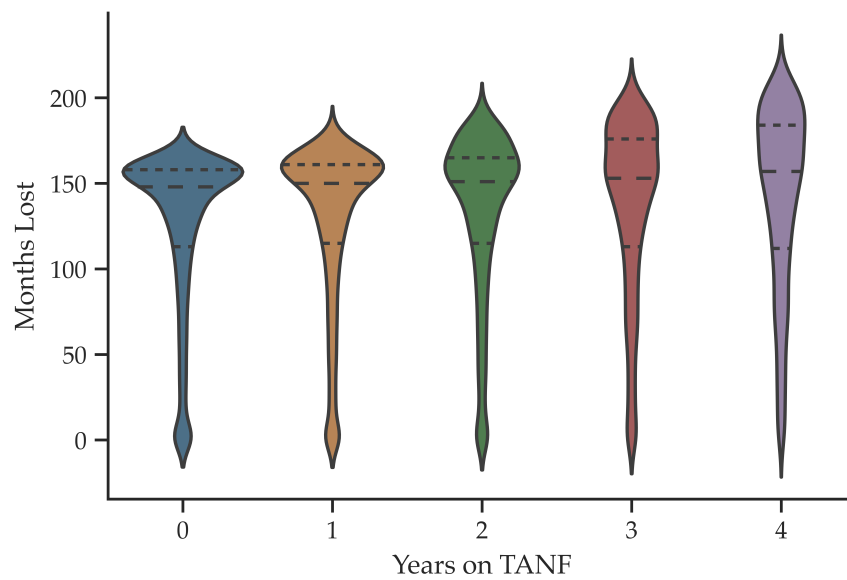
Notes: This figure shows the distribution of welfare utilization from October 2011 to December 2019 by time limit assignment separately for 5 groupings of families on welfare (0-1,1-2,...,4-5 years of welfare assigned as a result of the policy in 2011). Time limits are assigned based on prior use, so by definition those with 3 years assigned have already utilized 2 years of benefits. We convert the time used to a fraction by dividing the total number of months used by the total number of months assigned to them.

200 months of eligibility due to the policy change, which makes sense given that most families have children between 5-10.

A.3.2 Additional Results on Lost Eligibility

This appendix section provides additional specification and robustness for the empirical work done in section 4.2.2. For completeness, we begin by repeating the elasticity results in levels. These are shown in table A.3. As a robustness check on the main results provided in table 4, we add controls one by one to the main result shown in the first panel for the welfare eligibility and labor market outcomes. We summarize all results in table A.4. Results are largely unchanged between specifications. As mentioned in the main text, the fixed effects on the time limit granted on the date of the policy and the number of children are critical for these results, due to the fact that they directly impact the welfare assignment. The additional fixed effects just provide precision to the estimates.

Figure A.7: Distribution of Eligible Potential Months Lost By Time on Welfare



Notes: This figure shows the distribution of months of potential welfare use lost due to the imposition of the policy by years on welfare. The main text defines this value more clearly. The figure is broadest where there is the most mass in the distribution. The thicker dotted lines show the median, while the more frequently dotted lines on either end show the 25th and 75th percentiles of each respective distribution.

Table A.3: The Marginal Value of a Month of Welfare Eligibility (Levels)

		On TANF	Employed	Wages	SNAP	Medicaid
All Families	Months Lost	−0.02** (0.0)	0.07** (0.01)	43.85** (5.83)	0.0 (0.01)	−0.02** (0.01)
	Observations	17611	17611	17611	17611	17611
	Adj. R-squared	0.18	0.25	0.24	0.09	0.07
	F-statistic	496.96	4636.42	2092.43	129.89	173.47
Young Child	Months Lost	−0.04* (0.02)	0.29** (0.05)	190.72** (31.4)	−0.14** (0.04)	−0.08* (0.04)
	Observations	9920	9920	9920	9920	9920
	Adj. R-squared	0.2	0.25	0.25	0.11	0.08
	F-statistic	214.94	2193.81	1710.47	209.79	155.69
Older Child	Months Lost	−0.01* (0.01)	0.0 (0.02)	5.36 (11.06)	0.04* (0.02)	0.01 (0.01)
	Observations	7691	7691	7691	7691	7691
	Adj. R-squared	0.16	0.26	0.24	0.09	0.07
	F-statistic	159.12	1939.6	815.88	55.94	59.68

Notes: This table shows the results of the absorbing least squares regression with the explanatory variable the months lost on the date of the policy. The outcome variables are, respectively, cumulative TANF use for the subsequent 8 years, cumulative quarters employed and total quarterly wages for the subsequent 3 years, and SNAP and Medicaid enrollment for the subsequent 8 years after the policy. We use fixed effects to isolate the effect due to the lost months of eligibility. The fixed effects included are months of TANF assigned (60 minus months used), number of children, age in years, high school graduation status, race, and county of residence. All regressions control for employment, wages, and TANF use in the year leading up to the reform. The sample includes all women between the ages of 18 and 30. This restriction on ages accounts for the majority of the sample and isolates the women with the lowest barriers to employment and those with children of approximately the same age. The age cutoff for young child is 2 years of age, which is approximately the median child age. Standard errors, clustered by months of TANF use, are in parenthesis below each coefficient. ** Indicates p -value is less than .01, * Indicates p -value is less than .05, + Indicates p -value is less than .1.

Table A.4: The Marginal Value of a Month of Eligibility: Robustness

		(1)	(2)	(3)	(4)	(5)
Estimates	On TANF	−0.12** (0.03)	−0.22** (0.03)	−0.21** (0.03)	−0.18** (0.03)	−0.18** (0.03)
	Observations	16798	16798	16798	16798	16798
	Adj. R-squared	0.11	0.12	0.15	0.19	0.19
	F-statistic	534.1	539.13	571.91	424.67	452.19
	Employed	0.15** (0.03)	0.18** (0.03)	0.18** (0.03)	0.18** (0.03)	0.15** (0.03)
	Observations	14843	14843	14843	14843	14843
	Adj. R-squared	0.13	0.13	0.14	0.14	0.16
	F-statistic	1819.74	1719.69	1760.09	1794.69	1706.34
	Wages	0.2** (0.06)	0.38** (0.06)	0.38** (0.06)	0.37** (0.07)	0.31** (0.07)
	Observations	14843	14843	14843	14843	14843
	Adj. R-squared	0.14	0.15	0.15	0.15	0.17
	F-statistic	1557.47	1495.53	1507.92	1502.45	1424.08
Dummies Used	Time Limit	✓	✓	✓	✓	✓
	N. Children	✓	✓	✓	✓	✓
	Age		✓	✓	✓	✓
	Race			✓	✓	✓
	County				✓	✓
	Education					✓

Notes: This table shows the results of the absorbing least squares regression with the explanatory variable the months lost on the date of the policy. The outcome variables on the left hand side are, respectively, cumulative TANF use for the subsequent 8 years, and cumulative quarters employed and total quarterly wages for the subsequent 3 years. To see the impact of the fixed effects, we use fixed effects to isolate the effect due to adding the controls, we add them one-by-one beginning with the baseline of just number of children and time limit assigned. The fixed effects included are months of TANF assigned (60 minus months used), number of children, age in years, high school graduation status, race, and county of residence. All regressions control for employment, wages, and TANF use in the year leading up to the reform. Standard errors, clustered by months of TANF use, are in parenthesis below each coefficient. ** Indicates p -value is less than .01, * Indicates p -value is less than .05, + Indicates p -value is less than .1.