The Efficacy of Hiring Credits in Distressed Areas

Jorge Pérez Pérez* and Michael Suher†

November 2019

Abstract: We analyze the efficacy of hiring tax credits, particularly in distressed labor markets. These types of programs have proven hard to assess as their introduction at the state level tends to be endogenous to local conditions and future prospects. We conduct an empirical study of a hiring tax credit program implemented in North Carolina in the mid 1990s, which has a quasi-experimental design. Specifically, the 100 counties in the state are ranked each year by a formula trying to capture their economic distress level. The generosity of the tax credits jumps discontinuously at various ranking thresholds allowing for the use of regression discontinuity methods. Our estimates show fairly sizable and robust impacts on unemployment - a $9,000 credit leads to a nearly 0.5 percentage points reduction in the unemployment rate. The attendant increase in employment levels appears to be around 3%.

1 Introduction

Hiring tax credits are a commonly used tool at the state level and as part of federal programs to address both short-run downturns and longer run economic distress. They are place-based in that they aim to revitalize a specific geography rather than individual workers. Prior evidence on their efficacy has been mixed, with zero or small positive impacts on employment seen depending on the type of credit.

The empirical evaluation of these policies is difficult as their enactment is typically designed to be endogenous to expected economic prospects or local economic distress. The direction of the bias is also not clear. Significantly poor economic performance may swamp

*Banco de México (e-mail: jorgepp@banxico.org.mx)

†Federal Reserve Board (e-mail: michael.suher@frb.gov). We thank Matthew Baird, Patrick Button, Serena Canaan, Pierre Cahuc, John Friedman, Cecilia García, Daniel Hammermesh, David Jaume, Gustavo Leyva, Ann Battle Macheras, Diego Mayorga, Pascal Michaillat, Silda Nikaj, Matthew Turner, Gonzalo Vasquez-Bare, Cullen Wallace, and seminar participants at APPAM 2017, AREUEA 2018, Banco de México, the Fed Systems Committee Regional Analysis Conference, the Furman Center, the IZA Junior/Senior Symposium 2017, the NTA Annual Meeting 2016 and the UEA Meeting 2019 for valuable comments and suggestions. Pérez Pérez acknowledges funding from Fulbright-Colciencias. The views expressed are those of the authors and not necessarily those of the Federal Reserve Board or Banco de México.

estimates of program impact even if they are positive. Alternatively, natural mean reversion in areas which recently experienced negative shocks could be incorrectly attributed to the policy intervention. We examine a series of tax credit programs enacted in the state of North Carolina whose structure enables causal estimates of policy impacts. The programs assign credit size based on an economic distress rank built from several economic variables. They include thresholds at which credit size jumps discontinuously, allowing for regression discontinuity (RD) estimates. We find significant reductions in unemployment rates which grow over time, and are on the order of 0.5 percentage points for a $9,000 credit. We find employment increases of around 3%. Further, we show that difference-in-differences estimations would have failed to find an employment impact, indicating the importance of accounting for time varying unobservables.

Some prior papers have considered the theoretical case of subsidizing hiring. Neumark (2013) highlights the potential outsize benefit of policies like hiring credits aimed at stimulating labor demand during a recession, where downward wage rigidities will mute the impact of labor supply policies. Kline and Moretti (2013) augment a spatial equilibrium model with persistent long run differences in unemployment rates across areas, as is empirically observed. Firms in high unemployment/low productivity areas post too few vacancies due to excessive hiring costs providing a rationale for subsidizing hiring in distressed areas. Amior and Manning (2018) find that distressed areas experience serially correlated negative demand shocks that lead to swift and continual outflows of employment. Population leaves as well, but not as fast as employment, leading unemployment rates to remain elevated.

A known drawback of hiring credits, in theory and practice, is wastage, meaning a large share are claimed by firms whose hiring would have taken place absent the program. Bartik (2001) notes that credits are typically small enough though that they can still compare favorably to other employment subsidy policies even after accounting for the wastage. The use of credits during recessions and/or in distressed areas can reduce the concerns of wastage, though churning may still be an issue. This is where firms increase hiring while simultaneously increasing separations. Because our analysis focuses on overall employment levels and unemployment rates, estimates of program effects will be net of any churning. Further, the program we study required firms to maintain the increased payroll levels from new hires in subsequent years for the credits to fully pay out. Crowd-out of hiring in sectors not targeted for the subsidies may also be a concern, though Michaillat (2014) argues that in areas with high unemployment rates, an abundance of job seekers limits any increase in labor market tightness implying larger local job multipliers from subsidized hires.

Our paper contributes most directly to a literature which has sought to empirically measure if, and to what degree, hiring credits are effective. Neumark and Grijalva (2015) use cross-state variation in the adoption of hiring credits to estimate their impact. They find no impacts on employment growth in general but small positive effects during recessionary periods or when programs incorporate recapture provisions. They do not consider the size of the credits, though. In order to deal with the endogeneity of the adoption of these programs, they rely on projected counterfactual employment trends based on state industrial composition prior to enactment of a policy. While such predicted employment measures are known to correlate strongly with actual employment changes at decennial frequency, they may fail to capture shorter run counter-factual employment trends, particularly during recessions. Even though our use of a regression discontinuity design can arguably offer more causal estimates of program effects, we produce difference in differences estimates as well. This nests the paper better with prior attempts at evaluating hiring credits that rely on that technique, and allows us to characterize the direction and extent of endogeneity bias. We also extend the analysis by considering the impact of hiring credits on unemployment rates.

The federal Empowerment Zone program (EZ) in the U.S., which includes hiring credits, was studied by Busso et al. (2013) who identify program impacts using areas with rejected EZ applications or later round EZ designations as control areas. Freedman (2013) studies the similarly conceived Enterprise Zone program using a regression discontinuity design approach made possible by a specific implementation of the program in Texas relying on neighborhood level poverty rate cutoffs. Both studies find positive employment effects for residents of neighborhoods selected for the programs. The policy we investigate differs in that it targets a broader county rather than a specific neighborhood.

More recently, Cahuc et al. (2018) use difference-in-differences and IV strategies to evaluate a hiring credit program introduced in France during the Great Recession restricted to small firms and low-wage workers. They find significant impacts on employment at eligible firms. They also find the program to have been particularly cost effective, though demonstrate through simulations that this was dependent on the program being both temporary and unanticipated.

A potential source of bias in evaluations of hiring credits is analyzed by Chirinko and Wilson (2016) who exploit cross-state variation in hiring credit adoption with a focus on the potential for fiscal foresight, wherein programs that are pre-announced could lead firms to initially depress hiring and then ramp up once the credits become available. They find evidence of positive impacts on employment at a lag of two to three years, consistent with our findings. They also find pre-program dips, which can upwardly bias estimates of program effects by 33%. Our reliance on annual rather than monthly employment levels should help alleviate this bias.

We describe the mechanics of North Carolina's hiring tax credits in section 2. In section

3, we summarize our data sources and give an overview of the labor market during our sample period. We describe the estimation strategy in section 4. Section 5 presents the estimation results. Section 6 concludes.

2 North Carolina's Hiring Tax Credit Programs

In the mid-1980s, North Carolina government officials were concerned with the divergence in economic fortunes among the state's 100 counties. A tax incentive program began in 1988 to address the situation in the least economically robust counties. The state Department of Commerce was tasked with ranking counties by level of economic distress each year from 1 to 100 using a legislatively defined formula and inputs. The rankings were then used to segment counties into discrete tiers which determined the size of tax credits a county's could claim. Significant revisions to the program were introduced in 1996 and in 2007, though the basic framework of ranking calculation and tiers was maintained. The final iteration of the program ended in 2014 and was not replaced, though the county rankings continue to be computed (Program Evaluation Division, 2015). The tax credit size histories are summarized in Table 1.

Table 1: Credit Size by Distress Rank (Dollars)

	Distress									
Years	10	20	30	40	50	60	70	80	90	100
Wave 0: 1988-1995	2,800									
Wave 1: 1996-2006	12,500	3,000-4,000				500-1,000				
Wave 2: 2007-2013	12,500				5,000				750	

In this study we focus on the William S. Lee program which began in 1996, and is denoted as wave 1 in Table 1, as it provides the cleanest quasi-experimental set-up (the specifics of the wave 0 and wave 2 iterations of the tax credit program are detailed in the appendix). Unlike its predecessor, the Lee program extended eligibility to firms in all 100 counties, but continued to reserve the larger credits for firms in the more distressed counties. Credits of \$12,500 were available to the 10 most distressed counties designated as tier 1. Firms in moderately distressed tier 2 counties could receive credits between \$3,000 and \$4,000 per new hire, while those in the least distressed tier 3 could receive between \$500 and \$1,000.[1] Our

[1]Under the official program definitions, there are five tiers in wave one of the program. Tier 1 is the same as our definition. We combine tiers 2 and 3 together and tiers 4 and 5 together as they have similar program intensity as measured by the credit size for which they are eligible.

analysis will focus on the comparison between tier 1 and tier 2, and the average differential credit size of \$9,000.[2]

The Lee program specified county rankings based on three inputs: unemployment rates, income per capita, and population growth. The appendix Table A.1 shows an example of this process for one year of the program. Counties are first ranked separately, 100 to 1, on each input. These three sub-rankings are then summed and ordered to create the 1 to 100 distress rank for the coming year. The ten lowest ranked/most distressed counties are assigned to Tier 1. Counties ranked 11 or higher start being assigned to Tier 2. Over the course of this program, the number of counties designated tier 1 each year was increased as combinations of low population and high poverty rates were added as overrides to the distress ranking system.[3] By 2006, the final year of the program, 28 counties were designated tier 1 and eligible for the largest credits.

Because counties were re-ranked every year, treatment status was not always constant, with occasional slippage between tiers, in addition to the legislated expansion of the lowest tier over time. Figure 1 shows the geographic distribution of county tier designations for the first and final years of the William S. Lee program.[4]

Only firms in certain industries were eligible for the program, with the main ones being manufacturing, wholesale trade, warehousing, and those related to data processing. While there was no requirement that hires be of certain types of workers, such as those currently unemployed, they had to be new employees (*i.e.* not transferred from another area in the state), working full-time and paid above the county average wage. The size of the credit was based on the county where the workplace was located, not on an employee's county of residence, which could differ for some commuters. The credit paid out over four years with

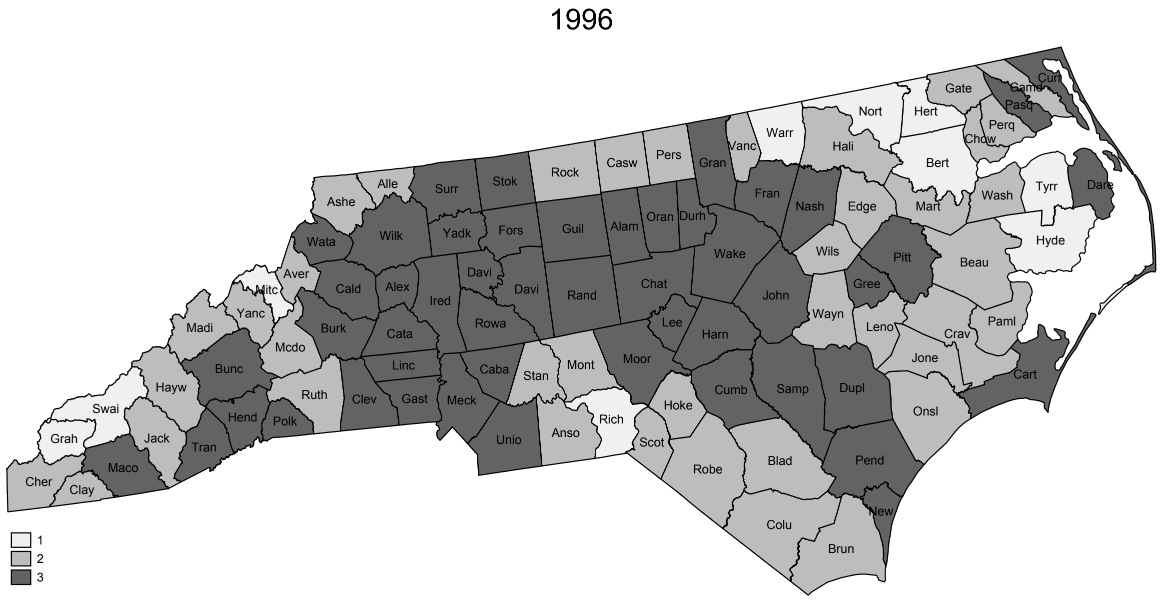
[2]This is the tier 1 \$12,500 minus the average credit of \$3,500 for tier 2. Similar RD estimates comparing tier 2 versus tier 3 are confounded by the prior iteration of the program. Even though the ranking procedure changed somewhat between wave 0 and wave 1, there is still a discontinuous change in the probability of wave 0 program eligibility at the rank 50 threshold separating tier 2 and tier 3.

[3]The overrides were as follows: a county's first time being designated tier 1 would persist through the subsequent year, even if its subsequent ranking would place it in a higher tier. Starting in 2000, counties with population below 10,000 and poverty rate above 16% were automatically tier 1, with the population threshold raised to 12,000 in 2002. Additionally, counties with population below 50,000 and poverty rate above 18% had their tier reduced by one from what their distress rank would otherwise dictate.

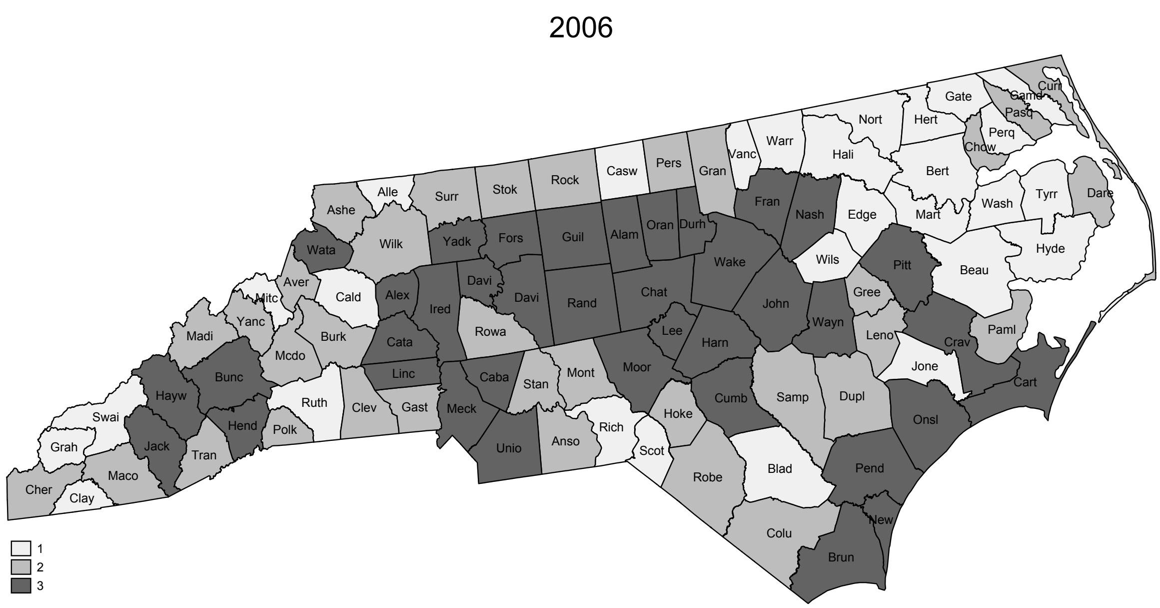
[4]One prominent feature of the map is the clustering of distressed counties in the eastern part of the state. These are the least urbanized counties and so tend to have lower income per capita and population growth, directly impacting their rankings. They also have low population levels, which, once the tier designation overrides for population and poverty levels are introduced in 2000, leads even more counties in this region to be designated into the most distressed tier 1 as seen in panel b of Figure 1. A North Carolina Department of Commerce assessment of the program (Fain (2001)) notes that this targeting was intentional in the design of the program. Plant closings in traditional North Carolina industries like textiles, apparel, furniture and tobacco manufacturing were having a disproportionately negative impact on the more rural and economically distressed parts of the state.

Figure 1: County Tier Status

(a)



(b)



Note: Tier status of the 100 North Carolina counties are presented for the first and final years of the William S. Lee program. Tier 1 counties are the most economically distressed.

latter year installments forfeited if the firm reduced its total number of employees.

Beyond hiring tax credits, the Lee program also offered other forms of incentives, most notably for investment in machinery and equipment (M&E) and research and development expenditures (R&D). These additional incentives had softer discontinuities in generosity at tier thresholds though and their benefits flowed primarily to the least distressed tier 3 counties, allowing us to largely isolate the impact of the hiring credits from these other incentives.[5]

Building 1 to 100 rankings using a somewhat ad-hoc choice of inputs meant high performing counties often received lower tier status than clearly more distressed counties. This became even more pronounced once small population overrides to the rankings were introduced. The same difference in rank can also be associated with different sized gaps in economic performance at different points in the ranking distribution. For example, the two counties ranked 10 and 30 can be quite different from each other, while the two counties ranked 30 and 50 are fairly similar to each other.

A more continuous and robust measure of distress was proposed by the state Department of Commerce though not adopted (Department of Commerce, 2014). Figure 2 shows unemployment, income per capita and population growth for the counties sorted by the distress rank of each input in 1996, and the tier to which each county was assigned. Two facts stand out. First, small differences in an input variable can lead to large differences in input specific sub-rankings. Second, counties with very similar inputs can end up in different tiers, through the effects of adding up the three sub-rankings and the overrides. Both overall distress rankings and input sub-rankings vary widely over time for each county. A county moves 6 positions in the overall distress ranking every year on average, with most of these shifts coming from changes in relative population growth and unemployment rate.

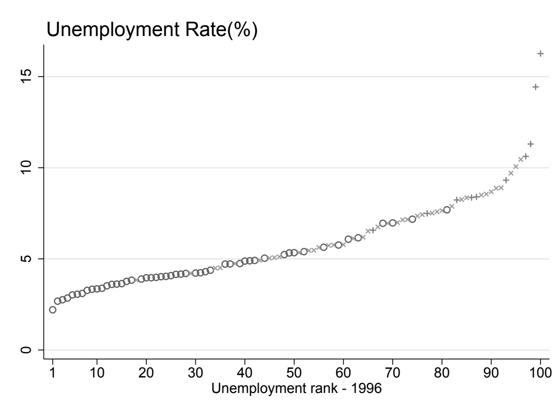
In section 4 we demonstrate that lower ranked counties do have lower population growth, higher poverty, and lower per capita income. There is no evidence of discontinuities in pre-treatment conditions at the program thresholds though. The ranking variable is also not strongly correlated with post-treatment outcomes after controlling for tier status, allowing us to do comparisons between counties farther away from the cut-offs.

Two potential concerns that could arise with the research design we implement are antic-

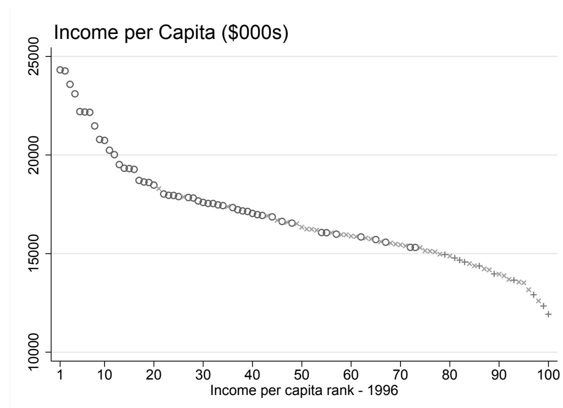
[5] The R&D credit size did not vary with the tier system. The M&E credit was constant across tiers but applied to any size investment by tier 1 firms but only investment in excess of $100-$200 thousand for tier 2 firms. Over three quarters of both M&E and R&D credits dollars went to firms in the least distressed tier 3 which we exclude from the analysis. The Lee program also included separate credits for job training and central administrative offices but these were minor with each accounting for less than 2% of total Lee Act credits generated (Fain, 2005). Cerqua and Pellegrini (2014) examine the effectiveness of these kind of capital subsidies in Italy. They exploit discontinuities in subsidy assignment through firm, finding that they are effective in boosting firm growth. We focus on hiring credits instead of capital subsidies here.

Figure 2: Distress Ranks per Input and Tier Designation

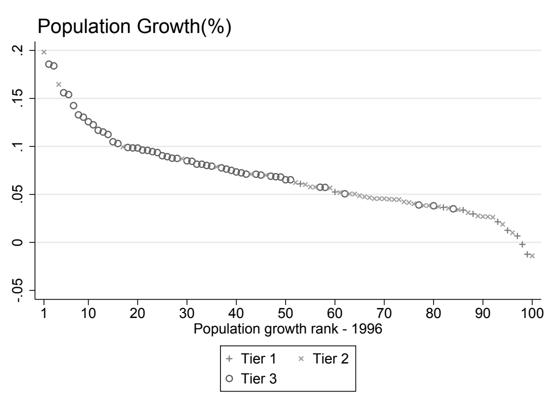
(a)



(b)



(c)



Note: County level economic indicators are arrayed by initial distress rank per input at the outset of the first wave of the program. Different symbols denote the different treatment tiers.

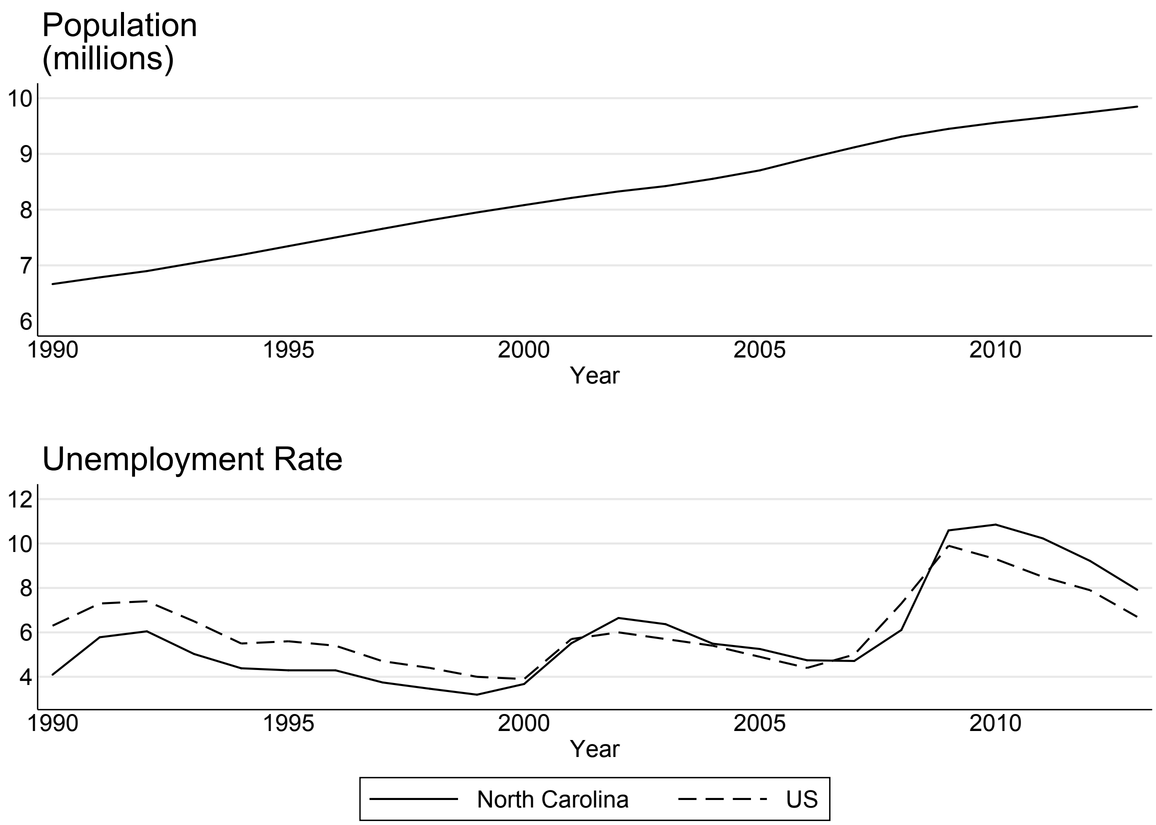
ipation effects and manipulation of program participation thresholds. As studied in Chirinko and Wilson (2016), firms may artificially depress current hiring if they anticipate becoming eligible for credits in the near future, which would lead to overestimates of true program impact. As mentioned above, we use annual rather than monthly data, meaning any distortion in hiring timing induced by the program would need to be over long timescales. Further, the initial enactment of the program occurred in the middle of 1996 and became effective immediately. Updated tier designations for each future year were not finalized and announced until December of the prior year further limiting the scope for anticipatory hiring delays.

Manipulation of treatment status around the eligibility threshold is always a concern with regression discontinuity designs. Because eligibility was determined from county level data collected by independent official sources, and was largely based on relative standing among all other counties, strategic manipulation does not seem possible. One area where some manipulation could possibly have occurred is in revisions to the program, beginning in 2000, which introduced overrides to the tier designation process based on absolute, rather than relative, population and poverty. It is plausible that county lobbying could have influenced the calibration of these overrides though the choice of round numbers for the thresholds is evidence against any such strategic behavior.

3 Data

Employment, unemployment, and labor force data is from the Bureau of Labor Statistics (BLS). Hiring and separations data come from the Census Bureau's Quarterly Workforce Indicators (QWI). Tier status comes from annual reports issued by the state of North Carolina and archived versions of the Commerce Department's website. The distress rankings were provided by the North Carolina Department of Commerce for some years and for other years reconstructed using data from the BLS, population and income data from the Bureau of Economic Analysis, and poverty data from the Census and the United States Department of Agriculture, in conjunction with the rules laid out in the legislation creating and amending the programs. Our sample period runs from 1990 to 2006. Figure 3 shows overall conditions in North Carolina's economy during the program period. The state's overall population is growing throughout, highlighting the extent to which the program was aimed at addressing significant within state regional divergence in economic performance. The state's unemployment rate tracks closely with the national business cycle and the program span includes the 2001 recession.

Figure 3: Population and Unemployment Rate in North Carolina



Note: Population data is from the Bureau of Economic Analysis. Unemployment data is from the Bureau of Labor Statistics.

4 Estimating the Effect of Hiring Credits

In this section we describe our estimation strategy. We lay out the difficulties of estimating the effect of hiring credits lay out a strategy to take advantage of the assignment of subsidies based on distress ranks.

North Carolina's 100 counties were assigned each year to three groups defined by the subsidy program tiers. We focus on the wave of the program running from 1996 to 2006. The most distressed counties are in tier 1 and receive the highest subsidy amount, $12,500. There are 10 counties designated tier 1 in the initial program year, 1996, but distress rank ties and amendments to the program in later years adding additional assignment rules leads this number to fluctuate between 10 and 28 in later program years. This is our treatment group. Tier 2 counties are the next most distressed and comprise our control group. Firms in those counties are eligible for hiring credits ranging from 3 to 4 thousand dollars. The least distressed counties are designated tier 3 and are eligible for hiring credits ranging from 500 to 1 thousand dollars. We estimate the effect of the program by comparing the evolution of employment and unemployment across counties in tiers 1 and 2. To avoid making comparisons between extremely different counties we exclude any counties ever designated as tier 3 as these contain major cities which may have very different dynamics compared to small distressed counties. The average subsidy for tier 2 counties is $3,500 compared to the $12,500 in tier 1 counties so our program effect estimates are for the $9,000 difference.

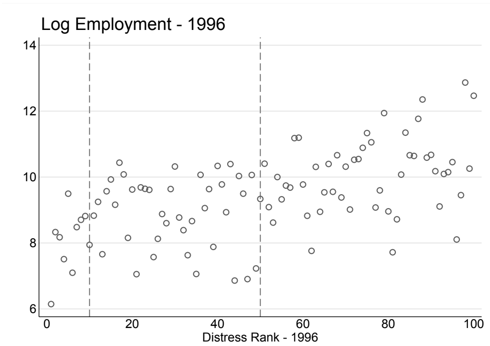
Figure 4 shows the relationship between the levels and changes of the outcome variables and the economic distress rank before the beginning of the program. Two messages emerge from this Figure. As expected, there is an overall negative relationship between economic outcomes and economic distress. Unemployment is higher for more distressed counties, while log employment is lower. However, this relationship is smooth across the tier cutoffs. Moreover, the slope of this relationship is small, suggesting that the economic distress rank is not strongly correlated with these outcomes, within each tier and across tiers for counties with similar ranks. The correlation between the distress rank and the change in the outcomes before the onset of the program is also weak, although the unemployment rate is more variable over time for more distressed, smaller counties.[6]

If the economic distress ranking were completely random, then counties would be assigned subsidy amounts randomly and we could compare counties across tiers. In practice, the distress rank is weakly correlated with economic variables. If counties that get assigned into tier 1 have systematically worse unobservables that imply different trajectories

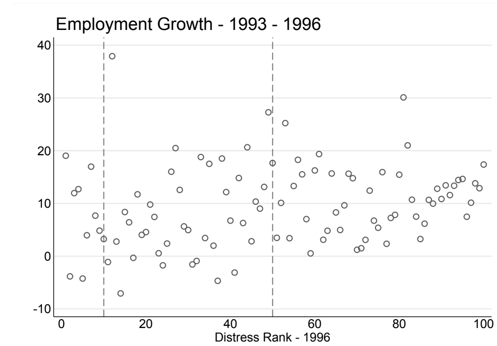
[6] In Appendix Figure B.1 we also show that the relationship between the distress rank inputs, and the variables that determine overrides to tier assignment, is smooth across tier cutoffs.

Figure 4: Relationship between the Distress Rank and Program Outcomes

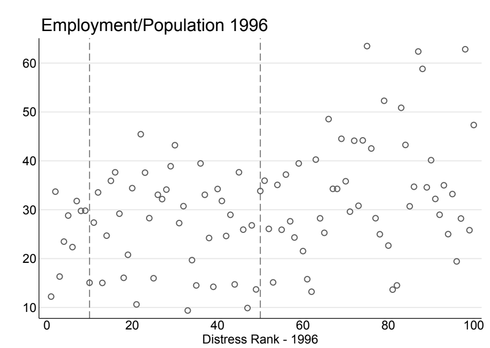
(a)



(b)

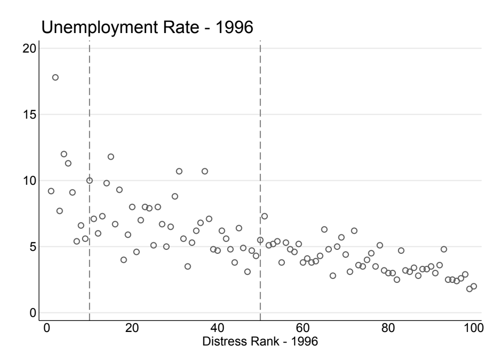


(c)

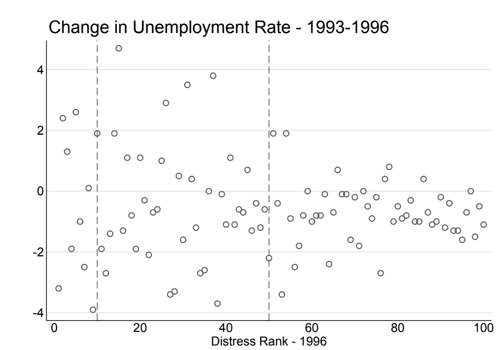


(d)

(e)



(f)



Note: County level economic indicators are arrayed by initial distress rank at the outset of the first wave of the program. Vertical lines denote the thresholds where credit size jumps discontinuously.

of employment and unemployment even in absence of the program, then estimates from a difference-in-differences approach will be biased.

We tackle this problem by exploiting the discontinuities in tier assignment based on the economic distress rank. As assignment to the program is redetermined each year, we follow Cellini et al. (2010) in implementing a dynamic regression discontinuity design estimation, that allows for contemporaneous as well as lagged effects of the program.

Our baseline specification is:

$$Y_{ctk} = \beta_0 + \gamma_c + \gamma_t + \gamma_k + \theta_k tier1_{c,t-k} + \nu_k f(rank_{c,t-k}) + \beta_k X_{c,t-k} + \varepsilon_{ctk}. \quad (1)$$

Here, Y_{ctk} is the outcome of interest for county c at time t measured k years after treatment designation. γ_c, γ_t, and γ_k are fixed effects for county, year, and time since treatment designation. $tier1_{c,t-k}$ is a dummy variable equal to 1 whenever a county is assigned to tier 1. The θ_k coefficients measure the program impact at various lags. $f(rank_{c,t-k})$ is a function of the county ranking at time $t-k$. The coefficients on the ranking and on the controls when included are allowed to vary with time since treatment assignment k.

Allowing for lagged effects is essential since the hiring subsidy programs may take a few years to gain traction and have a noticeable effect on employment (Neumark and Grijalva, 2015). Moreover, including lagged effects allows us to compare counties that have a similar history of treatment. By looking at the differences in these coefficients, we can assess how the effect of the program changes over time. We estimate this specification using only counties in tiers 1 and 2, from 1996 to 2006.

The RD approach allows us to obtain unbiased estimates of the effect of the program, as long as the conditional expectation of the unobservables that enter ε_{ctk} in (1) with regards to the county ranking vary smoothly across the tier 1 cutoff. Additionally, these estimates will address bias from mean reversion if the mean-reverting component of ε_{ctk} arising from transitory shocks does not change discretely across the cutoff. This happens despite the fact that tier 1 assignment depends indirectly on the outcomes at the time of program assignment through the ranking function (Chay et al., 2005).

There are some issues with implementing this specification in our setting. The first issue is that tier 1 assignment was not entirely based on the economic distress rank. Counties could not be redesignated out of tier 1 due to an improved distress rank until after two years. From 2000 onwards, high poverty and low population based rules are added as overrides to the formula for tier assignment. Wong et al. (2013) propose and assess methods for dealing with multiple assignment variables in an RD framework. They recommend excluding units who are assigned based on additional rules, and estimating equation (1) as a sharp discontinuity design using only counties assigned on the basis of the running variable being considered, in this case, the distress rank. Another approach is to classify the counties who change tiers because of these overrides as "defiers", and instrument tier 1 status with tier 1 assignment based on the distress rank as in a fuzzy discontinuity design. We adopt both strategies in section 5.

The second issue is the reduced sample size available to estimate each year's program effect. For the comparison between tier 1 and tier 2, we only have 70 counties available.[7] This limits our ability to estimate a large number of parameters or implement non-parametric estimators. We reduce the number of coefficients to estimate by making three assumptions: a constant treatment effect assumption and two assumptions on the conditional expectation of outcomes given the distress rank that seem consistent with data before the beginning of the program.

If the effect of the program is constant over time, we can take advantage of the repeated execution of the program. Our constant treatment effect assumption is that the effect of the program only depends on the number of years that have passed since the program takes place. For each program year and county, we generate a new set of observations for outcomes stretching from two years before to four years after treatment designation. These spans of observations are then pooled together to estimate a single panel regression. So for a given county, there are repeated, overlapping observation windows for [1993,1999], [1994,2000], [1995,2001], etc. for program designation rounds taking place in 1995, 1996, and 1997, etc., respectively. We account for the multiple appearances of a given county by year outcome by clustering standard errors by county.

Our additional assumptions concern the functional form of $f(rank_{c,t-k})$. Figure 4 suggests a linear conditional expectation function of changes in the outcomes given a distress rank. Moreover, the functional form of this relationship does not seem to change at the cutoff threshold. We therefore assume that $f(rank_{c,t})$ is linear and remains constant on either side of the assignment cutoff. We also try including the ranking input variables themselves, unemployment rate, income per capita, and population growth, as controls. These can help compensate for the low predictive power of the rank variable for future county outcomes.

To address concerns about this functional form assumption and about extrapolation far away from the cutoffs, we also calculate local estimates that only use variation near the cutoff. We pool changes in outcomes following each year of the program, and compare the means of these changes across tiers. We conduct hypothesis tests on these local estimates using randomization inference (Cattaneo et al., 2015, 2016).

[7]Tier 1 has counties ranked 10 and below and tier 2 has counties ranked between 10 and 50. More than 50 counties enter the regressions at some point both through ranking ties and ranking changes over time. Our parameter estimates only use counties in tier 1 and tier 2 every time.

We also experiment with how we account for the dynamic nature of the subsidy program. Consider the effect of the program two years after it is enacted in year t. In year $t+1$, the county would receive the contemporaneous and lagged effect of the program. If the county receives the subsidy in year $t+1$ as well, by year $t+2$ it would experience lagged effects of the program in t and $t+1$ together. Moreover, receiving the program in t may have altered the probability of receiving it in $t+1$. We include indicators for prior treatment status but not subsequent treatment. Cellini et al. (2010) show that in this setting, the estimated effects can be interpreted as "Intention to Treat" (ITT) effects, where employment outcomes are not affected only by the receipt of the subsidy but also by changes in the probability of receiving the subsidy in the future.

Cellini et al. (2010) also develop a "Treatment on the Treated" (TOT) estimator which accounts for the indirect impact of initial treatment on the probability of treatment in future years.[8] We apply their method with the following regression:

$$Y_{ct} = \beta_0+\gamma_c+\gamma_t+\sum_{k=0}^{K}(\alpha_k m_{c,t-k}+\theta_k tier1_{c,t-k}m_{c,t-k}+\nu_k m_{c,t-k}f(rank_{c,t-k})+\beta X_{c,t-k})+\varepsilon_{ct}. \quad (2)$$

Here, Y_{ct} is the outcome of interest for county c at time t. γ_c and γ_t are fixed effects for county and year. $tier1_{c,t-k}$ is a dummy variable equal to 1 whenever a county is assigned to tier 1. $m_{c,t-k}$ is a dummy for being in the 50 most distressed counties at time $t-k$. The θ_k coefficients measure the program impact at various lags. The coefficients on the ranking and on the controls when included are allowed to vary with time since treatment assignment k. We interact treatment assignment with the $m_{c,t-k}$ dummies to use only variation from the most distressed counties every year. By including the history of treatment assignment up to each outcome observation, θ_k will be a TOT estimate. This is meant to isolate the impact of having received treatment k years ago and not in subsequent years.

5 Results

We now turn to the regression discontinuity estimates. Figure 5 portrays graphical evidence. Counties are arrayed by initial distress rank relative to the threshold where credit size increases. A linear fit in county rank is included, which is constrained to have the same slope on either side of the threshold. Given the prior evidence that program effects appear

[8]The ITT and TOT terminology for the dynamic RD design is from Cellini et al. (2010). It should not be confused with the ITT and TOT terminology for randomized experiments with partial compliance. At the county level and accounting for the overrides, there is full compliance with the hiring credits program.

only with a lag, we focus here on three-year differences. Outcomes for counties entering the program at different points in time are pooled together here. Outcomes are weakly correlated with distress rank, with the assumption of a linear relationship with rank appearing reasonable.[9]

Table 2 presents the dynamic ITT estimates from equation (1). Rows 1 and 2 show estimates for log employment, which are progressively increasing for treated counties relative to the counter-factual through three years after treatment designation. The three-year effect with the ranking input variables as additional controls shows an employment increase of 3.6%. The same pattern of estimates is seen for the employment/population ratio in the next two rows. The effect three years after treatment is estimated to be around 1 percentage point. Rows 5 and 6 show that the unemployment rate in tier 1 counties is reduced relative to control counties two years after treatment and continues its relative decline through three years after. Relative to control counties, the unemployment rate in treated counties is between 0.5 and 0.7 percentage points lower after three years with the lower estimate the result of adding additional county level controls. For reference, over the course of the program, unemployment rates averaged 6.6% for the sample overall and 7.9% for the most distressed counties comprising the treatment group.

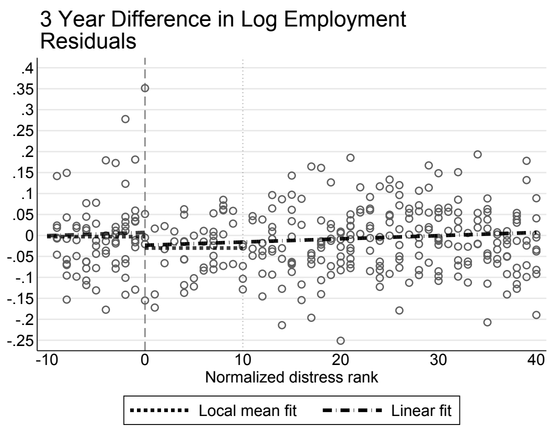
The presence of lagged effects can be further explained by the persistency of treatment status within counties. More than half of the treated counties receive tier 1 status for more than three years, and most counties receive three years of treatment conditional on receiving treatment a first time.

Table 3 implements equation (2) for the dynamic TOT estimates using the two different approaches proposed by Wong et al. (2013) for handling multiple assignment rules. The OLS rows implement a sharp RD design and exclude counties assigned by rules besides the distress rank from the treatment group. IV rows implement a fuzzy RD design, where we include all counties assigned to tier 1 by any rule, and instrument for treatment using the primary assignment rule based on distress rank. All rows include county level controls in addition to the running variable. For employment growth, the OLS estimates of treatment are insignificant. This likely reflects the more demanding nature of this estimation relative to the ITT specification above as more parameters are being estimated simultaneously. On the other hand, the IV estimates show sizable effects three years after treatment designation,

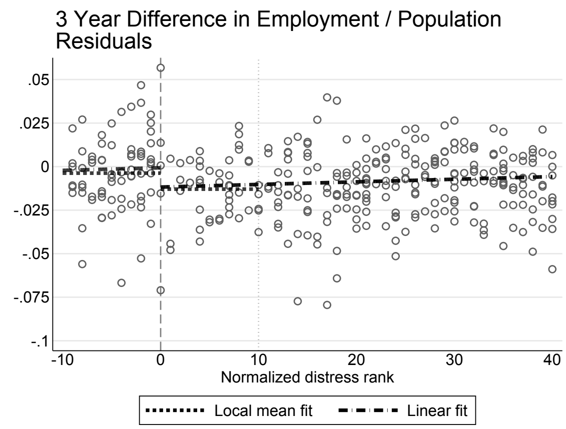
[9]There is an apparent outlier in panel (a) of Figure 5. This data point corresponds to Northampton county in 2005. It is plausible that this large employment effect is attributed to the program. The biggest Tier 1 credit recipient is Lowe's, and it has a distribution center in Garysburg in Northampton county. We do not have data on the jobs generate by Lowe's in every county, but aggregate data indicates that Lowe's generates 673 jobs in 2003, 271 in 2004, and 274 in 2005. This accounts for a total of 1,218 in direct hires generating credits, which is close to the the entire employment increase shown in the figure. In appendix table B.4 we show local estimates excluding this data point. The results are qualitatively similar.

Figure 5: Discontinuities in 3 Year Differences of Employment and Unemployment.

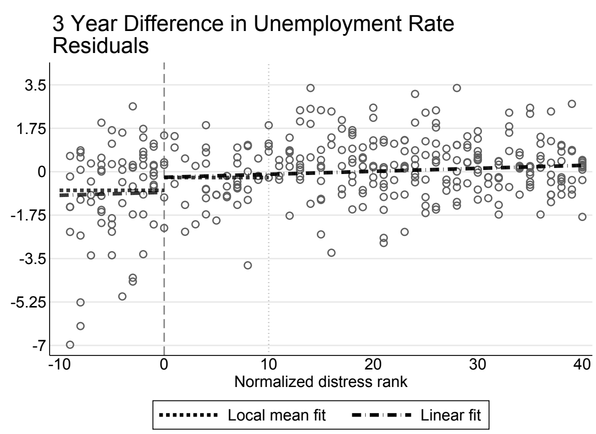
(a)



(b)



(c)



Note: Three-year differences in outcomes, 1996-2006. Sample mean plus residuals of a regression of the differenced outcomes on year dummies. Data is from the Bureau of Labor Statistics. Counties are arrayed by distress rank relative to the threshold. Counties to the left of the threshold are eligible for a larger hiring tax credit. The thicker lines are estimated linear control functions in distress rank. The thinner lines are means within a bandwidth of +/- 10 distress ranks.

Table 2: Regression Discontinuity ITT Estimates - Main Outcomes

	1 year later	2 years later	3 years later
Log Employment	0.005	0.013	0.030*
	(0.013)	(0.016)	(0.017)
with controls	0.006	0.016	0.036**
	(0.013)	(0.015)	(0.017)
Employment/Population	0.002	0.004	0.010*
	(0.004)	(0.005)	(0.005)
with controls	0.002	0.005	0.012**
	(0.004)	(0.005)	(0.005)
Unemployment Rate	0.075	-0.468	-0.748**
	(0.274)	(0.306)	(0.322)
with controls	0.188	-0.319	-0.507**
	(0.319)	(0.261)	(0.228)

Clustered standard errors in parentheses

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Note: N= 2,670. Each row comes from a separate estimation of equation (1) and reports the treatment effect estimates θ_k at one, two, and three years after treatment designation. Standard errors are clustered by county. All rows include fixed effects for year, county, time since tier designation, and prior treatment history, and a linear control function in the distress rank. Additional controls are the lagged three-year averages of the unemployment rate and real income per capita and population growth since the most recent census, which are the three inputs to the distress rank.

but a reduction in employment growth one year after the program starts. The employment to population results display a similar pattern, but the IV estimates do not show significant employment decreases. The unemployment results are clearer. Three years after, eligibility for the largest hiring credits has reduced the unemployment rate in a county by 0.5 percentage points according to the OLS specifications with defiers excluded. IV results for unemployment rate reductions using the fuzzy discontinuity approach are larger, around 1.0 to 1.2 percentage points.

A priori we expected TOT estimates to be larger than the ITT estimates as the beneficial effects of initial treatment on unemployment rates should reduce the probability of receiving the program in future years. This is the case for our employment estimates. For our unemployment rate estimates, the OLS ITT and TOT estimates are similar, but the IV estimates are larger. This may be due to the lower efficiency and higher finite sample bias of IV estimates. The fact that the OLS TOT estimates are largely similar to the OLS ITT estimates implies that the benefits of the program are not large enough to change the probability of future treatment. In other words, initially treated counties converge towards control counties but do not tend to overtake them in terms of economic performance, at least as measured by the distress ranking inputs of lagged averages of unemployment rates, income per capita, and population growth.

5.1 Local Estimates

The regression discontinuity design assumes random assignment of treatment at the policy threshold with a control function allowing for more distant observations to contribute to estimation of the treatment effect. This requires assumptions about the shape of that control function. Cattaneo et al. (2015) propose a non-parametric estimation technique which uses randomization inference in a small neighborhood around the threshold.

We implement their approach in Table 4, where we show estimates for the effect of the program three years ahead.[10] The size of the estimation window where the random assignment assumption is deemed most plausible is determined by a series of covariate balance tests for progressively larger bandwidths around the threshold. We use the lags of the unemployment rate, employment growth, and share college as the covariates. This yields a bandwidth of +/- 6 with differing unemployment rates prior to treatment responsible for the rejection of balance for larger windows. The treatment effect is then estimated as a simple difference in means, which is visualized in overlays to Figure 5. P-values are generated

[10]Full estimation results for one, two and three years after treatment designation are in Table B.3 of the appendix.

Table 3: Regression Discontinuity TOT Estimates - Main Outcomes

Dependent Variable - Method	1 year later	2years later	3 years later
Log Employment - OLS	-0.033*	-0.005	0.009
	(0.016)	(0.014)	(0.019)
Log Employment - IV	-0.065*	0.038	0.072**
	(0.038)	(0.031)	(0.031)
Employment/Population - OLS	-0.008*	-0.001	0.007
	(0.004)	(0.004)	(0.005)
Employment/Population - IV	-0.016	0.010	0.023**
	(0.011)	(0.008)	(0.009)
Unemployment Rate - OLS	0.156	-0.578*	-0.542*
	(0.337)	(0.339)	(0.307)
Unemployment Rate - IV	-0.130	-1.030**	-1.177*
	(0.622)	(0.496)	(0.610)

Clustered standard errors in parentheses

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Note: N=770. Each row comes from a separate estimation of equation (2) and reports the treatment effect estimates θ_k at one, two, and three years after treatment designation. Standard errors are clustered by county. All rows include lags of the unemployment rate, income per capita, and population growth as controls. OLS rows correspond to a sharp RD design, which excludes from estimation of the treatment effects any counties designated as Tier 1 by an assignment rule besides the primary one based on distressed rank. IV rows correspond to a fuzzy RD design, which labels all Tier 1 counties as being treated and instruments for treatment with the distressed rank assignment rule.

from repeated random resampling of the counties within the bandwidth to either side of the threshold. Outcomes are pooled across all years of the program after year effects have been partialled-out. We present estimates for the balanced window as well as for windows of 10 and 20 ranks for more precision in the estimates but with less claim to random assignment of treatment.

The estimate for log employment in row one of Table 4 implies that receiving the credits increases employment in a county by 5% after three years, which is in between the parametric ITT and TOT results. The estimates fall to 2.8% for a bandwidth of 10 ranks, and 2.5% for a bandwidth of 20 ranks, suggesting that the counties farther from the threshold have materially different counterfactual trajectories than those in the narrower window. The estimate for employment/population implies an increase of 1.4 percentage points, which decrease to about 1 percentage point with a larger bandwidth.

The estimate for unemployment in row one of Table 4 implies that receiving the credits decreases the unemployment rate in a county by 0.05 percentage points after three years. This effect is small and different from the parametric estimates above. Expanding the window leads to more precisely estimated effects of -0.5 to -0.9 percentage points, in line with the parametric estimates. This suggests that mean reversion in unemployment rates is upwardly biasing these estimates, particularly for the most severely distressed counties ranked four and below, and excluded from the narrowest bandwidth.

One concern with the local estimates, particularly for the smallest window, is the impact of the introduction of the population/poverty overrides to the ranking scheme that begin with the 2000 program year. Because population growth is an input to the distress rank, counties with the slowest growth are both more likely to get a lower distress rank and to stay below the population thresholds that entail an override designation to tier 1. As with the ITT estimates and OLS TOT estimates, we exclude these "defier" counties from the sample for the local estimations to maintain a sharp RD. Because a lower distress rank is correlated with the probability of being a defier, many of the lowest performing counties just to the right of the tier 1 threshold will be removed from the control group once the overrides kick in after 2000. This will tend to bias control group outcomes upwards. With a larger control group and the inclusion of covariates in the parametric estimations this is less of a concern. With the local estimations, the non-random augmentation of the control group is more problematic.

To address this concern, we present local estimates in the lower panel of Table 4 where the sample is restricted to 1996-2002 (*i.e.* the four cohorts from 1996-1999 observed three years later). Focusing on the 6 ranks window, log employment increases by 3.6% and employment to population by 1.4 percentage points, both in line with the parametric results. The unemployment rate falls by 0.4 percentage points, slightly smaller than the parametric estimates, but still economically meaningful.

5.2 Additional Outcomes

The tax credits were limited in the industries which were eligible, with the main sectors being manufacturing, warehousing, wholesale trade, and data processing.[11] Rows 1 and 2 of Table 5 present RD estimates of equations (1) and (2) for aggregate employment within the targeted industries and non-targeted industries separately. The point estimates for log employment are imprecise, but show substantial increases in target industry employment of 7-8% and no evidence of increased employment in non-target industries. The employment

[11] At the inception of the program in 1996, eligible industries, with NAICS codes in parentheses, were Manufacturing (31-33), Warehousing (493), Wholesale Trade (42), Research and Development (541710) and Data Processing (Computer Systems Design & Related Services (54151), Software Publishers (511210), Software Reproducing (334611), Data Processing Services (514210), and On-Line Information Services (514191)). Beginning in 1999, also made eligible were Air Courier Services (492110), Central Administrative Office (551114), Electronic Mail Order (454110), and Customer Service Center (561422).

Table 4: Local Estimates of Effect 3 Years after Treatment

Time Range	Window	Dependent Variable						N
		Log Employment		Employment Population		Unemployment Rate		
1996-2006	6 ranks	0.053^{**}	[0.018]	0.014^{**}	[0.013]	-0.050	[0.884]	69
1996-2006	10 ranks	0.028^{*}	[0.089]	0.010^{**}	[0.022]	-0.508	[0.128]	119
1996-2006	20 ranks	0.025^{**}	[0.042]	0.010^{***}	[0.002]	-0.908^{***}	[0.000]	186
1996-2002	6 ranks	0.036	[0.168]	0.014^{*}	[0.064]	-0.410	[0.549]	39
1996-2002	10 ranks	0.016	[0.390]	0.009^{*}	[0.067]	-1.162^{**}	[0.017]	69
1996-2002	20 ranks	0.006	[0.734]	0.008^{*}	[0.072]	-1.541^{***}	[0.000]	109

P-values from randomization inference with 1000 replications in brackets.

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Note: Difference in mean outcomes for treated and control counties, for different bandwidths of distress ranks around the policy threshold, three years after treatment.

to population estimates are in rows 3 and 4. These estimates account for the employment variability in small counties, showing more precise employment increases. Target industry employment to population ratio increases are statistically significant and around 1.5 percentage points.

Table 5: Regression Discontinuity ITT and TOT Estimates - Other Outcomes

	ITT - 3 years later	TOT - 3 years later
Log Employment Target Industries	0.066 (0.043)	0.081 (0.073)
Log Employment Non-Target Industries	0.004 (0.025)	-0.007 (0.034)
Employment/Population Target Industries	0.014*** (0.005)	0.017** (0.008)
Employment/Population Non-Target Industries	0.000 (0.004)	-0.000 (0.007)
Log Hires Annual Total	0.058 (0.039)	0.179** (0.087)
Log Separations Annual Total	0.037 (0.038)	0.134* (0.070)

Clustered standard errors in parentheses

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Note: Both columns report treatment effect estimates θ_k three years after treatment designation. Standard errors are clustered by county. All the estimations include as controls lagged three-year averages of the unemployment rate and real income per capita and population growth since the most recent census, which are the three inputs to the distress rank. The ITT column reports estimates from equation (1). N=2,779 for hires and separations. N=2,700 for industry employment due to censoring of industry level data in some small counties. All rows include fixed effects for year, county, time since tier designation, and prior treatment history, and a linear control function in the distress rank. The TOT column reports IV estimates from equation (2). The IV estimates correspond to a fuzzy RD design, which labels all Tier 1 counties as being treated and instruments for treatment with the distressed rank assignment rule. N=770 for hires and separations. N=748 for industry employment due to censoring of industry level data in some small counties.

In Figure 6, we present the effect by each two-digit industry separately, estimating the effect on the ratio of each industry's employment to the overall county population. The three main target industries are at the top. The effects are small and insignificant for all the industries except manufacturing, which was the largest targeted industry. The estimated increase in the industry employment/population ratio is about 1.5 percentage points, similar to the ones estimated in Table 5 for all the targeted industries combined. The lack of discernible impact for warehousing may be due to its combination with a non-targeted industry, transportation. With both wholesale trade and warehousing, another factor potentially limiting tax credit use were the wage requirements. Credits could only be claimed for jobs paying above the county average wage which is more common in manufacturing than in the other two target industries.

As a robustness check, we re-estimate the RD specifications using Census QWI data on hires and separations. The employment gains we attribute to a hiring tax credit program should be stemming from increased hiring and not from decreased separations. Rows 5 and 6 of Table 5 present the RD estimates for hires and separations following a county's eligibility for the program. Though noisy, the estimates are consistent with employment gains resulting from increased hiring rather than decreased separations. In fact, separations show evidence of increasing as well. This tracks with the empirical observation in the labor literature that growing firms continue to have positive separation rates.

5.3 Difference in Differences Estimates

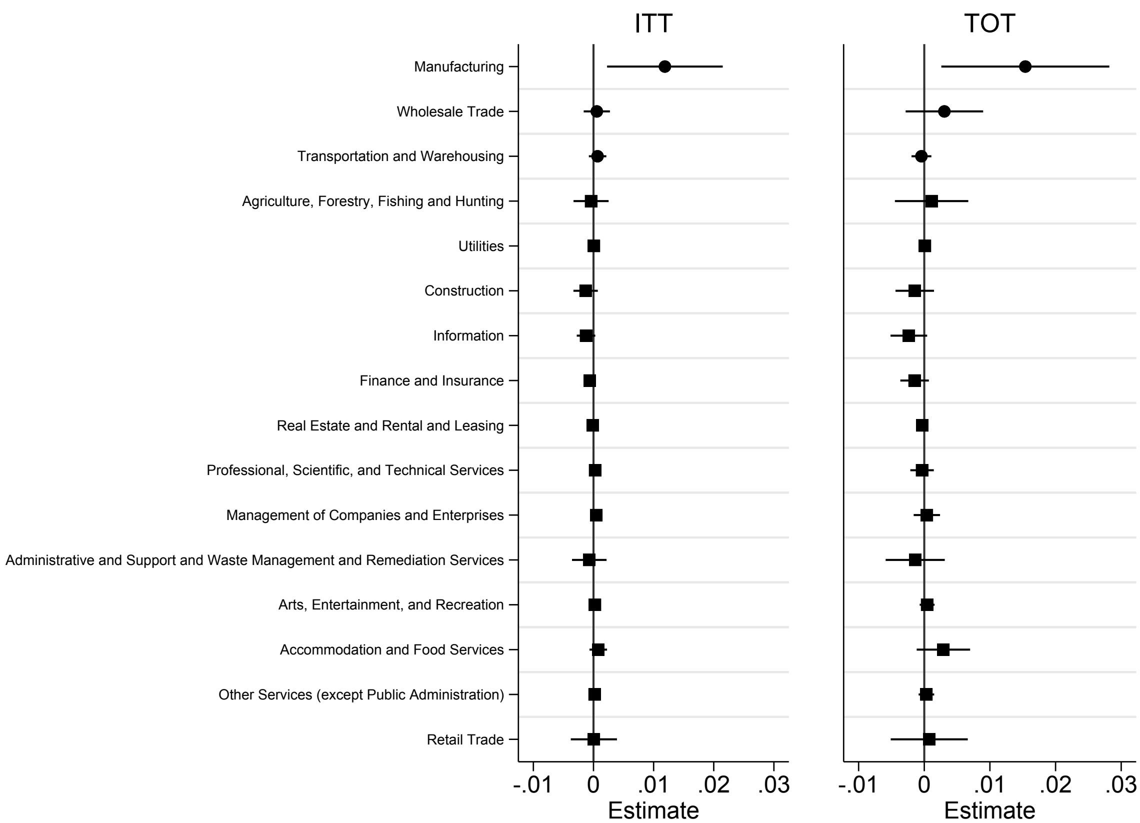
Our estimate of an employment impact of the hiring credits of around 3% is in some contrast to with prior studies of hiring credits. These have tended to find increases of 1%, and in some cases, no impact at all. Two possibilities for this discrepancy are methodology and setting. The program we study targets severely distressed areas where, as discussed above, hiring credits can theoretically be more impactful than in the average area. Besides the setting, prior studies have relied on difference-in-differences techniques which have the potential to be biased (positively or negatively) relative to a regression discontinuity approach. To assess the extent to which methodological differences are underpinning our larger impact estimates, we perform difference-in-differences (DD) estimation on the program.

Our basic specification is:

$$Y_{ct} = \beta_0 + \gamma_c + \gamma_t + \sum_{k=0}^{K} \theta_k tier1_{t-k} + \beta X_{ct} + \varepsilon_{ct}. \tag{3}$$

Here, Y_{ct} is the outcome of interest for county c at time t. γ_c and γ_t are county and year effects intended to capture permanent differences across counties and common shocks that affect all the counties in each year. $tier1_{t-k}$ is a dummy variable equal to 1 whenever a county is assigned to tier 1. The coefficients of interest, θ_k capture the contemporaneous and lagged effects of tier 1 status on the outcome variables.

Figure 6: Estimates of Effects on Industry Employment/County Population. 3 Years after Treatment



Note: Each row comes from separate estimations of equations (1) and (2) and reports the treatment effect estimates θ_k three years after treatment designation. Circle markers denote target industries. Square markers denote non-target industries. Health Care and Mining are excluded due to small sample sizes. For the TOT estimates, IV estimates are reported. The IV estimates correspond to a fuzzy RD design, which labels all Tier 1 counties as being treated and instruments for treatment with the distressed rank assignment rule. The bars around each coefficient are confidence intervals at the 95% level. Standard errors are clustered by county. Additional controls are the lagged three-year averages of the unemployment rate and real income per capita and population growth since the most recent census, which are the three inputs to the distress rank.

This difference in differences strategy is valid only if unobservables have a similar evolution over time across tiers. We also include control variables X_{ct} to allow for some county heterogeneity. We include lags of population growth, real income per capita, and the unemployment rate which comprise the inputs to the distress rankings. This addresses the possibility of counties evolving heterogeneously because of different initial conditions before the beginning of the program. It also addresses mean reversion in outcomes (Heckman et al., 1999). We also allow for county-specific linear time-trends.

In Table 6 we present the DD results with full controls three years after program assignment for comparison with our RD results from above. Log employment and employment to population show no impact, while the unemployment rate shows a reduction of around 0.5 percentage points. The full estimates are in appendix table B.2.

Table 6: Difference in Difference Estimates - Main Outcomes

Log Employment	Employment/ Population	Unemployment Rate
0.001	-0.001	-0.466**
(0.013)	(0.003)	(0.179)

Clustered standard errors in parentheses
* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Note: Sample size is 546 observations for 42 counties. The table depicts treatment effect estimates from of equation (3), three years after treatment designation. All rows include county effects, time effects, linear time trends interacted with county dummies, and lagged three-year averages of the unemployment rate, real income per capita and population growth since the most recent census, which are the three inputs to the distress rank. Standard errors are clustered by county to allow for serial correlation in the error term ε_{ct} within counties (Bertrand et al., 2004). P-values are calculated using a wild cluster bootstrap (Cameron and Miller, 2015) to account for the small number of clusters. We report clustered standard errors and significance tests from the bootstrap p-values with 500 replications. Full estimation results are in Table B.2 of the appendix.

The positive and relatively large RD estimates of a positive employment effect we present above imply that the difference in differences estimates of a null program effect are materially downward biased. This is consistent with the control function plotted in panel a of Figure 5 which is slightly upward sloping: within the treatment or control group, employment in higher ranked/less distressed counties tended to grow faster. In contrast, the control function for unemployment rates shows some tendency for convergence between more and less distressed counties independent of the program. In line with this, the difference in differences unemployment estimates are not notably biased relative to the RD unemployment estimates.

5.4 Take-up and program costs

An evaluation of incentive programs was produced by Lane and Jolley (2009) for the North Carolina General Assembly shortly after the William S. Lee program had ended and morphed into the similarly conceived Article 3J program. In its entirety, \$2.091 billion in tax credits were generated through the program from its 11 year run of 1996 to 2006. They estimate that 35% of these credits would never be used. This could result from future net job losses at incented firms leading to clawback of previously generated credit value. Insufficient future tax liability could also be a factor as credits taken could not exceed 50% of a firm's annual tax liability and unused credits could only be carried-forward for five years. The focus of this paper, the job creation credits, comprised 17% of the total amount of generated credits.[12] Though the program was much more generous on a per credit basis for the most distressed tier 1 counties, firms in those counties received only 14% of generated credit dollars, with 25% going to tier 2 counties, and over half of all credit dollars going to firms in the least distressed tier 3 counties.[13] Putting these figures together, we estimate that firms in our treated tier 1 counties took job creation hiring credits of \$56.6 million (or \$5.1 million per year) for 6,751 hires.[14]

In tier 1, on average about 24 firms per year generated some hiring credits for the 2001 to 2006 period when granular data is available. There are a number of factors in the design of the program that may have contributed to lower take-up rates. First, a firm had to operate in a target industry and have at least five existing full-time employees. It also had to provide health insurance for full-time positions, and not have recently violated environmental standards or safety requirements. The new hire needed to be for a full-time job and pay above the county average wage. Finally, the business would need sufficiently positive tax liability as credits could not exceed 50% of a taxpayer's total annual corporate income and franchise tax liability. A survey of North Carolina businesses which found that incentives ranked as only the 12th most important factor in company location decisions, behind things like access to skilled labor, highway access, tax rates, and regulator climate. In fact, 62% of surveyed executives at incented companies were unaware their company received an incentive, though it is not reported how many of these firm's were in the least distressed tier where credits per

[12] The breakdown of credits generated over the life of the William S. Lee program was job creation (17%), machinery and equipment investment (66%), R&D (15%), training (1%), and central administrative office (1%).

[13] This breakdown was based on program years 2002 to 2006.

[14] We arrive at these estimates for tier 1 as follows: We have only aggregated figures for 1996 to 2000, so we multiply the total credit generated figure for 1996 to 2000, \$1.064 billion, by the 0.17 share for job creation hiring credits, and then by 0.14 for the tier 1 share, and then by 0.65 for generated credits not taken. We add to this amount the values from more granular data on credits generated available for 2001 to 2006 that is also multiplied by 0.65 for generated credits not taken, and divide by 11 for the per year figure.

hire were only \$500 (Lane and Jolley, 2009).

The pairing of the hiring credits with other incentives for investment and R&D, while not atypical, mean our estimates are full program effects rather than pure hiring credit impacts. Because the discontinuities are less pronounced across tier and take-up of the other incentives is low in the counties on which we focus, we believe we are largely isolating a hiring credit effect.

The issue of wastage is also complicated in the specific program we study. The tax expenditures flowing to the most distressed counties were relatively small given the program impacts we find. Overall tax expenditures under the program were much larger though, as a large majority of benefits went to firms in economically robust counties where similar hiring and investment would likely have occurred absent the program. While in theory a program could be designed to completely exclude more economically robust areas from receiving expenditures, political limitations may make that implausible in practice.

5.5 Validity

While our use of an RD strategy should provide more confidence in the internal validity of our estimates, some bias is still possible. Because subsidies are place of work, spillover effects from any cross-county commuting induced by the program are possible. In instances where tier 2 counties border tier 1 counties this would entail overestimates of employment impacts and underestimates of unemployment impacts.

There are a number of aspects specific to the program under study that may limit the external validity of the results. Most important is the targeting of the credits to areas experiencing long-run economic distress. We would not expect the same size credit to have as large an impact in the average area that we find they have in distressed areas. Even in distressed areas, we cannot credibly project the impact of a much smaller per job credit as the benefit per dollar may be non-linear. Further, we do not estimate the impact of some versus no credits, but rather a big versus a small credit. However, under the assumption of a decreasing marginal product of labor, we would expect the former comparison to yield impacts at least as large as what we find.

6 Discussion

Hiring tax credits are a popular tool implemented at various times in many U.S. states as a way to lure businesses or revitalize moribund local economies. Assessing their efficacy is challenging though as their implementation is typically expressly endogenous to local conditions and expected future prospects. We make use of the unusual institutional features of a program in the state of North Carolina to get causal estimates of the impact of hiring tax credits on employment and unemployment rates. Our RD ITT estimates for employment are noisy but generally show a boost from the program of around 3%. In contrast, difference in differences estimates show no impact on employment, in line with prior studies of hiring credits. The evident downward bias in these latter estimates relative to the more credible RD estimates highlights the importance of accounting for time-varying unobservables impacting counterfactual treated county performance. We find substantial impacts on unemployment, with treated counties, those whose firms were eligible for large hiring tax credits, experiencing about 0.5 percentage points lower unemployment rates than under a counterfactual program offering much smaller credits.

References

Amior, Michael and Alan Manning (2018), "The Persistence of Local Joblessness." *American Economic Review*, 108, 1942–70.

Bartik, Timothy J. (2001), *Jobs for the Poor: Can Labor Demand Policies Help?* W.E. Upjohn Institute for Employment Research.

Bertrand, Marianne, Esther Duflo, and Sendhil Mullainathan (2004), "How Much Should We Trust Differences-in-Differences Estimates?" *The Quarterly Journal of Economics*, 119, 249–275.

Busso, Matias, Jesse Gregory, and Patrick Kline (2013), "Assessing the Incidence and Efficiency of a Prominent Place Based Policy." *American Economic Review*, 103, 897–947.

Cahuc, Pierre, Stéphane Carcillo, and Thomas Le Barbanchon (2018), "The Effectiveness of Hiring Credits." *The Review of Economic Studies*, Forthcoming.

Cameron, Collin A. and Douglas L. Miller (2015), "A Practitioner's Guide to Cluster-Robust Inference." *Journal of Human Resources*, 50, 317–372.

Cattaneo, Matias D, Brigham Frandsen, and Rocıo Titiunik (2015), "Randomization Inference in the Regression Discontinuity Design: An Application to Party Advantages in the US Senate." *Journal of Causal Inference*, 3, 1–24.

Cattaneo, Matias D., Rocio Titiunik, and Gonzalo Vazquez-Bare (2016), "Inference in Regression Discontinuity Designs Under Local Randomization." *Stata Journal*, 16, 331–367.

Cellini, Stephanie Riegg, Fernando Ferreira, and Jesse Rothstein (2010), "The Value of School Facility Investments: Evidence from a Dynamic Regression Discontinuity Design." *The Quarterly Journal of Economics*, 125, 215–261.

Cerqua, Augusto and Guido Pellegrini (2014), "Do Subsidies to Private Capital Boost Firms' Growth? A Multiple Regression Discontinuity Design Approach." *Journal of Public Economics*, 109, 114–126.

Chay, Kenneth Y., Patrick J. McEwan, and Miguel Urquiola (2005), "The Central Role of Noise in Evaluating Interventions That Use Test Scores to Rank Schools." *The American Economic Review*, 95, 1237–1258.

Chirinko, Robert S. and Daniel J. Wilson (2016), "Job Creation Tax Credits, Fiscal Foresight, and Job Growth: Evidence from U.S. States." CESifo Working Paper Series 5771, CESifo Group Munich.

Department of Commerce (2014), "Measuring Economic Distress in North Carolina." Technical report.

Fain, James T. (2005), "William S. Lee Act: 2005 Assessment of Results." Technical Report dp1357, State of North Carolina Department of Commerce: Division of Policy, Research & Strategic Planning.

Freedman, Matthew (2013), "Targeted Business Incentives and Local Labor Markets." *Journal of Human Resources*, 48, 311–344.

Heckman, James J., Robert J. Lalonde, and Jeffrey A. Smith (1999), "The Economics and Econometrics of Active Labor Market Programs." In *Handbook of Labor Economics* (O. Ashenfelter and D. Card, eds.), volume 3, chapter 31, 1865–2097, Elsevier.

Kline, Patrick and Enrico Moretti (2013), "Place Based Policies with Unemployment." *American Economic Review*, 103, 238–43.

Lane, Brent and G. Jason Jolley (2009), "An Evaluation of North Carolina's Economic Development Incentive Programs: Final Report." Technical report, Chapel Hill: University of North Carolina at Chapel Hill.

Michaillat, Pascal (2014), "A Theory of Countercyclical Government Multiplier." *American Economic Journal: Macroeconomics*, 6, 190–217.

Neumark, David (2013), "Spurring Job Creation in Response to Severe Recessions: Reconsidering Hiring Credits." *Journal of Policy Analysis and Management*, 32, 142–171.

Neumark, David and Diego Grijalva (2015), “The Employment Effects of State Hiring Credits.” IZA Discussion Papers 9146, Institute for the Study of Labor (IZA).

Program Evaluation Division (2015), “Final report to the joint legislative program evaluation oversight committee.” Technical Report 2015-11, North Carolina General Assembly.

Wong, Vivian C, Peter M Steiner, and Thomas D Cook (2013), “Analyzing regression-discontinuity designs with multiple assignment variables: A comparative study of four estimation methods.” *Journal of Educational and Behavioral Statistics*, 38, 107–141.

Online Appendix - Not For Publication

A Additional program details

The first iteration of North Carolina's non-discretionary tax incentive program, denoted as wave 0 in Table 1, began in 1988. The state Department of Commerce was tasked with ranking counties each year from 1 to 100 based on economic distress, which legislation defined as the combination of a high unemployment rate and low per capita incomes. Specifically, counties were ranked separately by unemployment rate and income per capita and then the two rankings were summed to produce an overall distress rank (potentially including ties). Businesses in the 20 most distressed counties were eligible for a $2,800 tax credit for each new full time employee hired. The number of eligible counties was progressively increased, and had reached 50 by the time the program ended in 1995. We were unable to find data on this program's implementation beyond the identity of the 50 counties eligible for credits in its final year which we use as a control variable.

A revamped program was launched in 1996 known as Article 3A or the William S. Lee program which is the focus of our study and referred to as wave 1. It continued the use of a county ranking scheme and added population growth as an input to the ranking process. It extended tax credits to all counties which varied in size based on groups of counties known as tiers. The largest credits of $12,500 were available to the 10 most distressed counties designated as tier 1. Firms in less distressed counties could receive credits between $500 and $4,000 per new hire. Over the course of this program, the number of counties eligible for the largest credit size was increased as low population and high poverty rates were added as overrides of the distress ranking system, with 28 counties designated tier 1 and eligible for the largest credits by the final year of the program in 2006.

The William S. Lee program was itself replaced in 2007 by the Article 3J program referred to as wave 2. This latter program operated in similar fashion to its predecessors, but with some changes to the credit eligibility formulas. Tier 1 - the most distressed - expanded to contain 40 counties eligible for credits of $12,500. The next most distressed 40 counties were in tier 2 and could receive $5,000 credits and the highest performing 20 counties in tier 3 could receive $750 credits. The distress ranking formula was amended to incorporate property value per capita alongside unemployment rate, income, and population growth. In 2014, the Article 3J program ended and was not replaced (Program Evaluation Division, 2015).

Table A.1: Distress Ranking Example - 2005

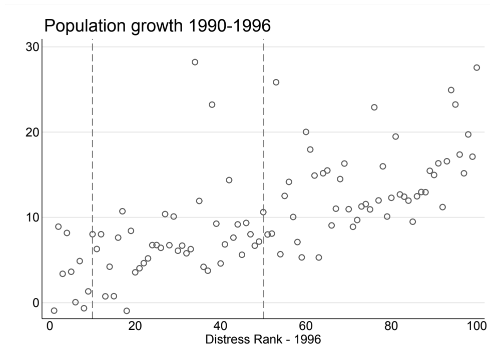
County	Population		Income		Unemp.		Distress		Pop.	Poverty	Override	Tier
	growth	rank	per cap	rank	rate	rank	sum	rank				
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Vance	-0.54%	96	$21,697	74	12.43	100	270	1	44,134	20.50	0	1
Halifax	-0.27%	86	$20,132	91	8.83	90	267	2	56,533	23.90	0	1
Scotland	-0.26%	85	$21,083	82	10.96	99	266	3	35,089	20.56	0	1
Hyde	-2.14%	100	$19,694	93	7.16	70	263	4	5,854	15.44	0	1
Washington	-0.85%	98	$20,926	85	7.69	78	261	5	13,507	21.76	0	1
Edgecombe	-1.39%	99	$22,373	64	9.59	96	259	6	55,583	19.59	0	1
Richmond	-0.21%	84	$21,195	81	9.37	93	258	7	46,053	19.56	0	1
Martin	-0.71%	97	$21,520	77	8.09	83	257	8	24,940	20.19	0	1
Warren	0.36%	62	$17,947	99	9.38	94	255	9	20,252	19.45	0	1
Yancey	-0.04%	80	$19,621	94	7.86	80	254	10	17,546	15.76	0	1
Mitchell	-0.09%	82	$20,004	92	7.59	77	251	11	15,770	13.83	0	2
Bertie	0.13%	76	$20,695	87	8.50	88	251	11	20,013	23.46	1	1
Anson	-0.35%	91	$22,536	60	10.25	97	248	13	25,690	17.77	0	2
Caswell	0.08%	78	$21,537	76	8.37	86	240	14	23,738	14.40	0	2
Robeson	0.65%	53	$18,238	98	8.31	85	236	15	125,394	22.81	0	2
North Carolina	1.33%		$27,644		5.47				8,418,493	12.23		

Note: Example of the computation of the distress ranking and tier assignment for the 15 most distressed in 2005 out of the 100 counties in North Carolina. The three inputs to the rankings are recent county population growth, income per capita, and unemployment rates. These are ranked separately from 100 to 1 in columns (2), (4) and (6) and then summed and ordered for the overall distress rank in column (8). The ten lowest ranked/most distressed counties are assigned to Tier 1. Counties ranked 11 or higher start being assigned to Tier 2, unless they trip an override based on a combination of low population and a high poverty rate that re-designates them as Tier 1. Overall values for North Carolina's 100 counties at the bottom.

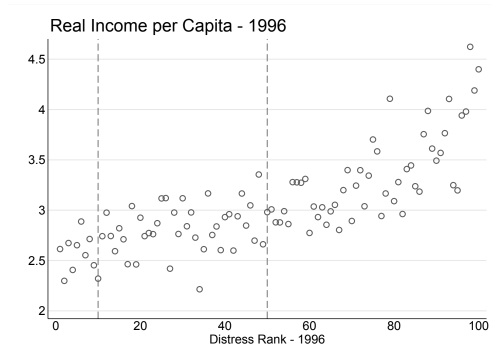
B Additional tables and figures

Figure B.1: Relationship between the Distress Rank and Ranking Inputs

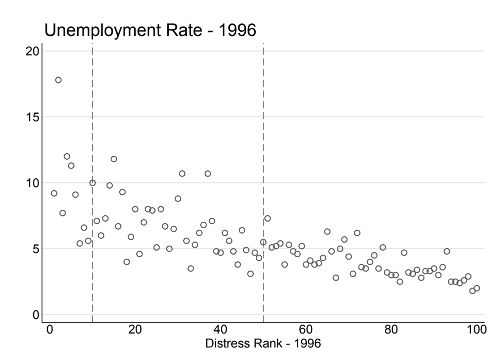
(a)



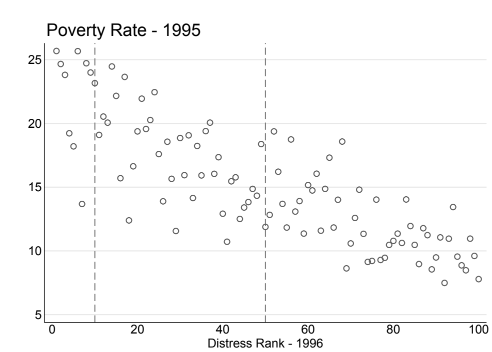
(b)



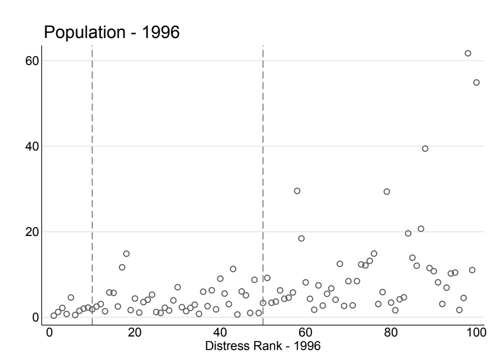
(c)



(d)



(e)



Note: County level economic indicators are arrayed by initial distress rank at the outset of the first wave of the program. Vertical lines denote the thresholds where credit size jumps discontinuously.

Table B.2: Difference in Differences Estimates

Dependent Variable	Log Employment		Employment/ Population		Unemployment	
	(1)	(2)	(3)	(4)	(5)	(6)
Tier 1	-0.007	0.001	-0.001	0.001	0.072	0.077
	(0.009)	(0.008)	(0.002)	(0.002)	(0.165)	(0.160)
Lag Tier 1	-0.021**	-0.016*	-0.004*	-0.004*	-0.121	-0.051
	(0.009)	(0.009)	(0.002)	(0.002)	(0.169)	(0.186)
Lag 2 Tier 1	-0.014	-0.011	-0.003	-0.002	-0.293**	-0.246**
	(0.009)	(0.008)	(0.002)	(0.002)	(0.117)	(0.112)
Lag 3 Tier 1	-0.002	0.001	-0.001	-0.001	-0.518***	-0.466**
	(0.014)	(0.013)	(0.003)	(0.003)	(0.189)	(0.179)
Lag 4 Population growth		-0.000		-0.001		0.042
		(0.003)		(0.001)		(0.030)
Lag 4 Real Income per capita		0.020		-0.003		0.655
		(0.045)		(0.012)		(0.694)
Lag 4 Unemployment Rate		-0.006**		-0.001		-0.111**
		(0.003)		(0.001)		(0.046)
R^2	0.998	0.998	0.981	0.982	0.810	0.826
N	588	546	588	546	588	546
Counties	42	42	42	42	42	42
County FE	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes
County trends	Yes	Yes	Yes	Yes	Yes	Yes

Clustered standard errors in parentheses

* $p < 0.1$ ** $p < 0.05$ *** $p < 0.01$

Note: Difference in differences estimates of equation (3). Standard errors are clustered by county. P-values for significance tests are calculated using a wild cluster bootstrap with 500 replications to account for the small number of counties. All columns include county and time effects, and linear time trends interacted with county dummies.

Table B.3: Local Estimates: 1,2 and 3 Years after Treatment

Window	Years	Dependent Variable			N
		Log Employment	Employment/ Population	Unemployment Rate	
6 ranks	1 Year	-0.001	0.000	-0.198	88
		[0.906]	[0.874]	[0.393]	
	2 Years	0.026	0.007	-0.287	78
		[0.157]	[0.126]	[0.392]	
	3 Years	0.053**	0.014**	-0.050	69
		[0.018]	[0.013]	[0.884]	
10 ranks	1 Year	-0.006	-0.000	-0.135	152
		[0.501]	[0.949]	[0.436]	
	2 Years	0.010	0.004	-0.495*	135
		[0.438]	[0.179]	[0.066]	
	3 Years	0.028*	0.010**	-0.508	119
		[0.089]	[0.022]	[0.128]	
20 ranks	1 Year	-0.004	0.000	-0.245**	236
		[0.530]	[0.757]	[0.046]	
	2 Years	0.006	0.004*	-0.643***	211
		[0.556]	[0.088]	[0.000]	
	3 Years	0.025**	0.010***	-0.908***	186
		[0.042]	[0.002]	[0.000]	

P-values from randomization inference with 1000 replications in brackets.

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Note: Difference in mean outcomes for treated and control counties, for different bandwidths of distress ranks around the policy threshold.

Table B.4: Local Estimates of Effect 3 Years after Treatment, Excluding Northampton county in 2005

Time Range	Window	Dependent Variable						N
		Log Employment		Employment Population		Unemployment Rate		
1996-2006	6 ranks	0.046**	[0.030]	0.013**	[0.026]	-0.043	[0.908]	68
1996-2006	10 ranks	0.023	[0.120]	0.009**	[0.031]	-0.507	[0.128]	118
1996-2006	20 ranks	0.020*	[0.086]	0.009***	[0.005]	-0.907***	[0.000]	185

P-values from randomization inference with 1000 replications in brackets.

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Note: Difference in mean outcomes for treated and control counties, for different bandwidths of distress ranks around the policy threshold, three years after treatment. Estimates exclude Northampton county in 2005.