

The Efficacy of Hiring Credits in Distressed Areas

Jorge Pérez Pérez* and Michael Suher†

November 2020

Abstract: We analyze the efficacy of hiring tax credits, particularly in distressed labor markets. These programs have proven hard to assess as their introduction at the state level tends to be endogenous to local conditions and 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 state’s 100 counties are ranked each year by a formula trying to capture their economic distress level. The generosity of the tax credits has discrete jumps at various ranking thresholds allowing for the use of regression discontinuity methods. Our estimates show sizable and robust impacts on unemployment - a \$9,000 credit leads to a nearly 0.5 percentage point reduction in the unemployment rate in the counties where the hiring credits were available. The attendant increase in employment levels appears to be around 3%.

1 Introduction

Hiring tax credits are a commonly used tool to address both short-run downturns and longer-run economic distress. The empirical evaluation of these policies is difficult as their enactment is endogenous to expected economic prospects or local labor market distress. The direction of the bias is also not clear. Poor economic performance may swamp estimates of program impacts even if they are positive. Alternatively, natural mean reversion — in areas that recently experienced adverse shocks — could be incorrectly attributed to the policy intervention.

*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 Michailat, Silda Nikaj, Ben Ranish, 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.

Earlier evidence had questioned hiring credit efficacy, often finding zero impacts on employment. In contrast, [Neumark and Grijalva \(2015\)](#) finds positive effects of state-level programs on employment during recessionary periods in the US. Similarly, [Cahuc et al. \(2019\)](#) sees benefits during the Great Recession of a hiring credit program targeted to small firms and low-wage workers in France. We build on this work by evaluating place-based hiring credits aimed at addressing regional disparities in a setting that permits causal estimates. We find significant reductions in unemployment rates and increases in employment in the distressed areas targeted by the hiring program.

We examine a program in the state of North Carolina that assigns county-level hiring credit eligibility based on an economic distress rank. The county rankings include thresholds at which credit size jumps discontinuously. Because program eligibility changes from year to year, we implement a dynamic regression discontinuity design following [Cellini et al. \(2010\)](#). We complement these findings with local non-parametric estimates that provide consistent evidence without a specific control function assumption. In treated counties, we find significant reductions in unemployment rates that grow over time and are on the order of 0.5 percentage points for a \$9,000 credit. We also find employment increases of around 3%, which we document are driven by increased hiring, as separation rates also rise. Further, employment impacts are concentrated in the manufacturing and wholesale trade sectors that were the program’s primary target. Finally, we produce difference-in-differences estimates to better nest the paper with prior attempts at evaluating hiring credits that rely on that technique and to allow us to characterize the direction and extent of endogeneity bias. Difference-in-differences estimations would have failed to find an employment impact, indicating the importance of accounting for time-varying unobservables.

Hiring credits—long a common policy tool—were often viewed dimly by economists due to concerns around inefficient distortion of the location of economic activity ([Kline and Moretti, 2014](#)). More recently, a theoretical case for the efficacy of hiring credits in some circumstances has begun developing as spatial inequalities and increasing home attachment have become better understood. [Neumark \(2013\)](#) highlights the potential outside 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 employment outflows. Population leaves as well, but not as fast as employment, leading

unemployment rates to remain elevated. [Coate and Mangum \(2019\)](#) document an increase in workers' home attachment across US labor markets, which translates into declining labor mobility and more persistent unemployment differences across places. [Zabek \(2019\)](#) models how residents' local ties, and outsiders limited interest in moving in, imply that place-based policies targeting depressed areas can transfer income with limited geographic distortion.

A known drawback of non-discretionary hiring credits — those made generally available to firms — is wastage, meaning that firms whose hiring would have taken place absent the program claim a large share of the credits. [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. For comparison, [Slattery \(2020\)](#) finds discretionary subsidies - those that target a specific firm - average around \$100,000 per job promised. Restricting the use of credits to recessionary periods or distressed areas can also limit wastage as non-subsidized hiring activity is low, a phenomenon our findings support.

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. The reduction of marginal production costs coming from the lower cost of new hires may also translate into higher business activity and higher employment levels by firms ([Saez et al., 2019](#)). While we do not see evidence of crowd-out from the hiring program, there are also no detectable spillover benefits to non-targeted sectors.

Our paper contributes most directly to a literature that has sought to empirically measure if, to what degree, and in what circumstances hiring credits are effective. Credible research designs are needed to address the inherent endogeneity of hiring program adoption and eligibility with current and expected economic performance. [Neumark and Grijalva \(2015\)](#) use cross-state variation in hiring credit programs to estimate their impact along with counterfactual employment trends based on state industrial composition. They find no impacts on employment growth in general but positive effects during recessionary periods and when programs incorporate recapture provisions. [Cahuc et al. \(2019\)](#) 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. [Sestito and Viviano \(2018\)](#) evaluate hiring credits in Italy during 2015 that targeted workers without permanent contracts. They find increased hiring for this group of workers in the aftermath of the policy relative to non-eligible workers.

Unlike these interventions, the program we investigate targets specific distressed labor

markets. Prior work on place-based hiring subsidies has focused on neighborhood-level programs in urban areas. Most have failed to detect beneficial effects though there are some exceptions (Neumark (2020)). The federal Empowerment Zone program (EZ) in the US, 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. In contrast, Neumark and Young (2019) use propensity score matching to evaluate both state enterprise and federal empowerment zone programs. They find evidence of reduced unemployment rates but no increases in employment or poverty reductions with their preferred empirical strategy.

Chirinko and Wilson (2016) analyze fiscal foresight, another potential source of bias in evaluations of hiring credits, wherein programs that are pre-announced could lead firms to initially depress hiring and then ramp up once the credits become available. Exploiting cross-state variation in hiring credit adoption, they report 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. Last, section 6 offers some discussion of our findings.

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 the level of economic distress each year from 1 to 100 using a legislatively defined formula and inputs. The state government used the rankings to segment counties into discrete tiers that determined the size of tax credits a county could claim. The program experienced significant revisions in 1996 and 2007, but they maintained the basic ranking calculation and tiers framework. The program’s final iteration ended in 2014, but the Department of Commerce continued computing the county rankings

(Program Evaluation Division, 2015). Table 1 summarizes the tax credit size histories.

Table 1: Credit Size by Distress Rank (Dollars)

Years	Distress Rank									
	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	

Note: Lower ranked counties are more economically distressed. Source: North Carolina Department of Commerce.

In this study, we focus on the William S. Lee program that began in 1996, which we denote as wave 1 in Table 1, as it provides the cleanest quasi-experimental set-up.¹ 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 ten 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, and those in the least distressed Tier 3 could receive between \$500 and \$1,000.² Our analysis will focus on the comparison between Tier 1 and Tier 2, and the average differential credit size of \$9,000.³

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, 1 to 100, on each input. These three sub-rankings are summed and ordered to create the 1 to 100 distress rank for the coming year. The ten lowest ranked, most distressed counties place in Tier 1. Counties ranked 11 or higher start in Tier 2. Throughout the program, the number of counties designated Tier 1 each year increased because of the addition of low population and high poverty rates as overrides to the distress ranking system.⁴ Twenty-eight counties were in Tier 1 and eligible for the largest credits by 2006, the final program year.

¹The appendix provides details about wave 0 and wave 2 of the tax credit program.

²Under 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 and tiers 4 and 5 together as they have similar program intensity as measured by the credit size for which they are eligible.

³This is the Tier 1 \$12,500 minus the average credit of \$3,500 for Tier 2. The prior iteration of the program is a confound for comparing counties between Tier 2 and Tier 3. 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.

⁴The overrides were as follows: a county's first-time designation to 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 a population below 10,000 and a poverty rate above 16% automatically placed in Tier 1. The population

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 a map with the geographic distribution of county tier designations for the first and final years of the William S. Lee program.⁵

Only firms in specific industries were eligible for the program, with the main ones being manufacturing, wholesale trade, warehousing, and those related to data processing. While hiring certain types of workers such as those currently unemployed was not required, they had to be new employees (*i.e.*, not intra-firm transfers from another area in the state), working full-time and paid above the county average wage. The workplace county determined the credit size, which did not depend on an employee’s county of residence. The credit paid out over four years with later installments forfeited if the firm reduced its total number of employees.

Beyond hiring tax credits, the Lee program 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. Their benefits flowed primarily to the least distressed Tier 3 counties, allowing us to isolate the hiring credits’ impact from these other incentives.⁶

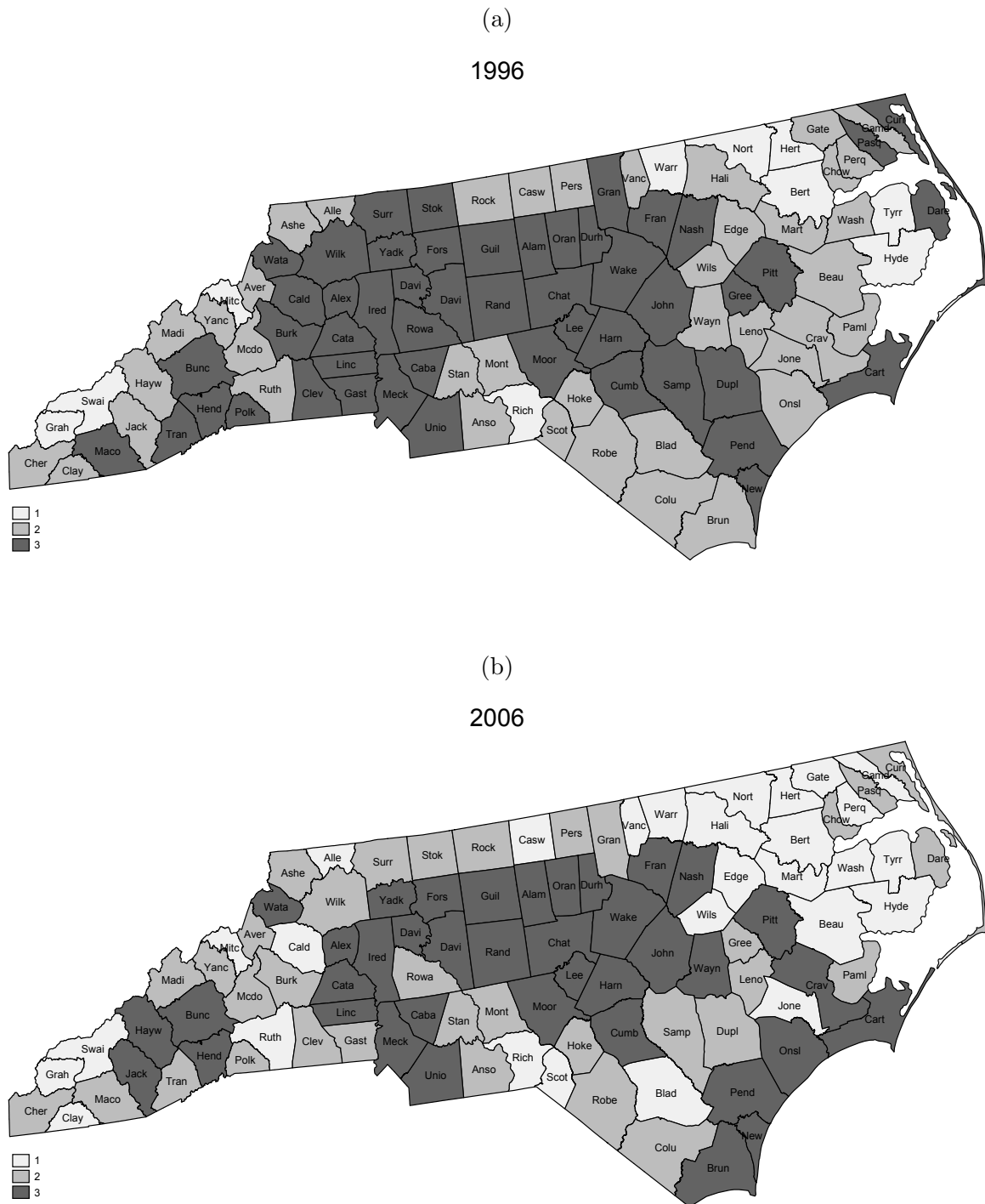
Building 1 to 100 rankings using a somewhat ad-hoc choice of inputs meant high performing counties often received a lower tier status than clearly more distressed counties. This ad-hoc nature became even more pronounced once the small population overrides to the rankings were introduced. Similar differences in rank are associated with different sized gaps in economic performance at various 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

threshold increased to 12,000 in 2002. Additionally, counties with a population below 50,000 and poverty rate above 18% had their tier reduced by one from what their distress rank would otherwise dictate.

⁵One 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 tend to have lower income per capita and population growth, directly impacting their rankings. They also have low population levels. Once the tier designation overrides for population and poverty levels are introduced in 2000, they lead 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 program’s design. Plant closings in traditional North Carolina industries like textiles, apparel, furniture, and tobacco manufacturing were having a disproportionately negative impact on the state’s more rural and economically distressed parts.

⁶The 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 above \$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. However, 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 kinds of capital subsidies in Italy. They exploit discontinuities in subsidy assignments across firms, finding that they are effective in boosting firm growth. We focus on hiring credits instead of capital subsidies here.

Figure 1: County Tier Status



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.

ranked 30 and 50 are similar to each other. The state Department of Commerce proposed a more continuous and robust measure of distress though it was not adopted. ([Department of Commerce, 2014](#)).

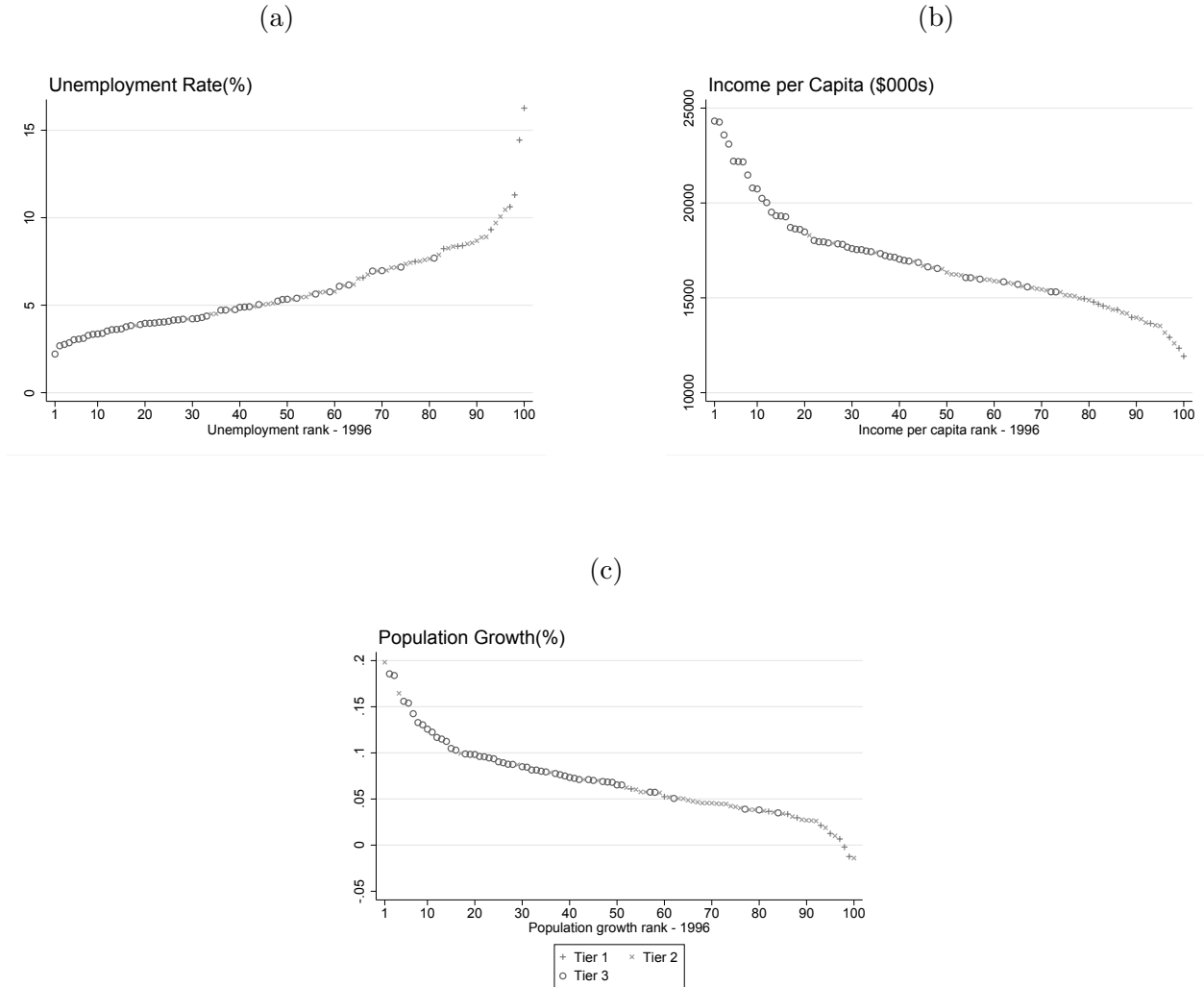
Figure 2 shows unemployment, income per capita, and population growth for the counties sorted by the distress rank of each input in 1996 and each county’s assigned tier. Two facts stand out. First, small differences in an input variable can lead to large differences in input specific sub-rankings. Second, 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 six positions in the overall distress ranking every year on average, with most of these shifts coming from in relative population growth and unemployment rate changes.

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

Two potential concerns that could arise with the research design we implement are anticipation 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 soon, which would lead to overestimating the program’s 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, limiting the scope for anticipatory hiring delays.

Manipulation of treatment status around the eligibility threshold is always a concern with regression discontinuity designs. Because the determination of eligibility used county-level data collected by independent official sources and largely depended on relative standing among all other counties, strategic manipulation does not seem possible. One area where some manipulation could 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.

Figure 2: Distress Ranks per Input and Tier Designation

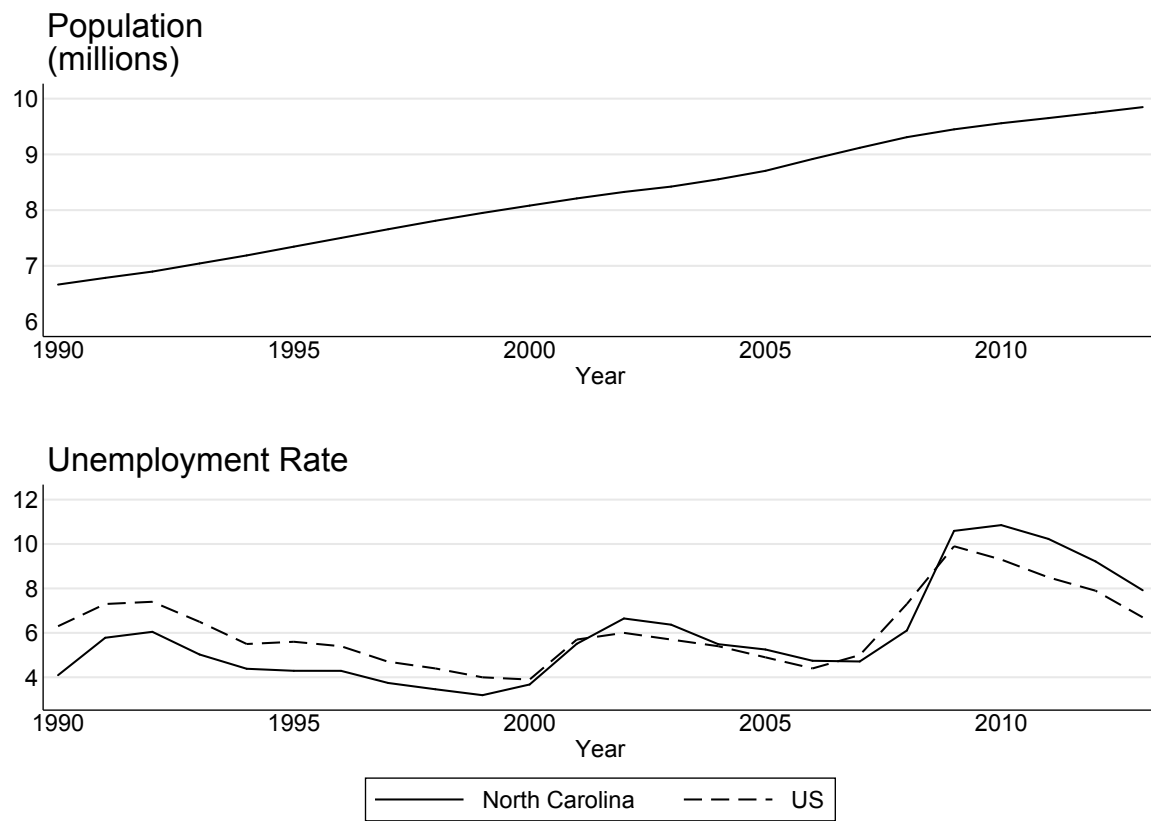


Note: County level economic indicators are arrayed by initial distress rank per input at the outset of the first wave of the program. Different symbols denote the different treatment tiers.

3 Data

Employment, unemployment, and labor force data are 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 come from the North Carolina Department of Commerce for some years. For other years they were 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 the overall conditions in North Carolina’s economy during the program period. The state’s total population is growing throughout, highlighting the extent to which the program aimed to address 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

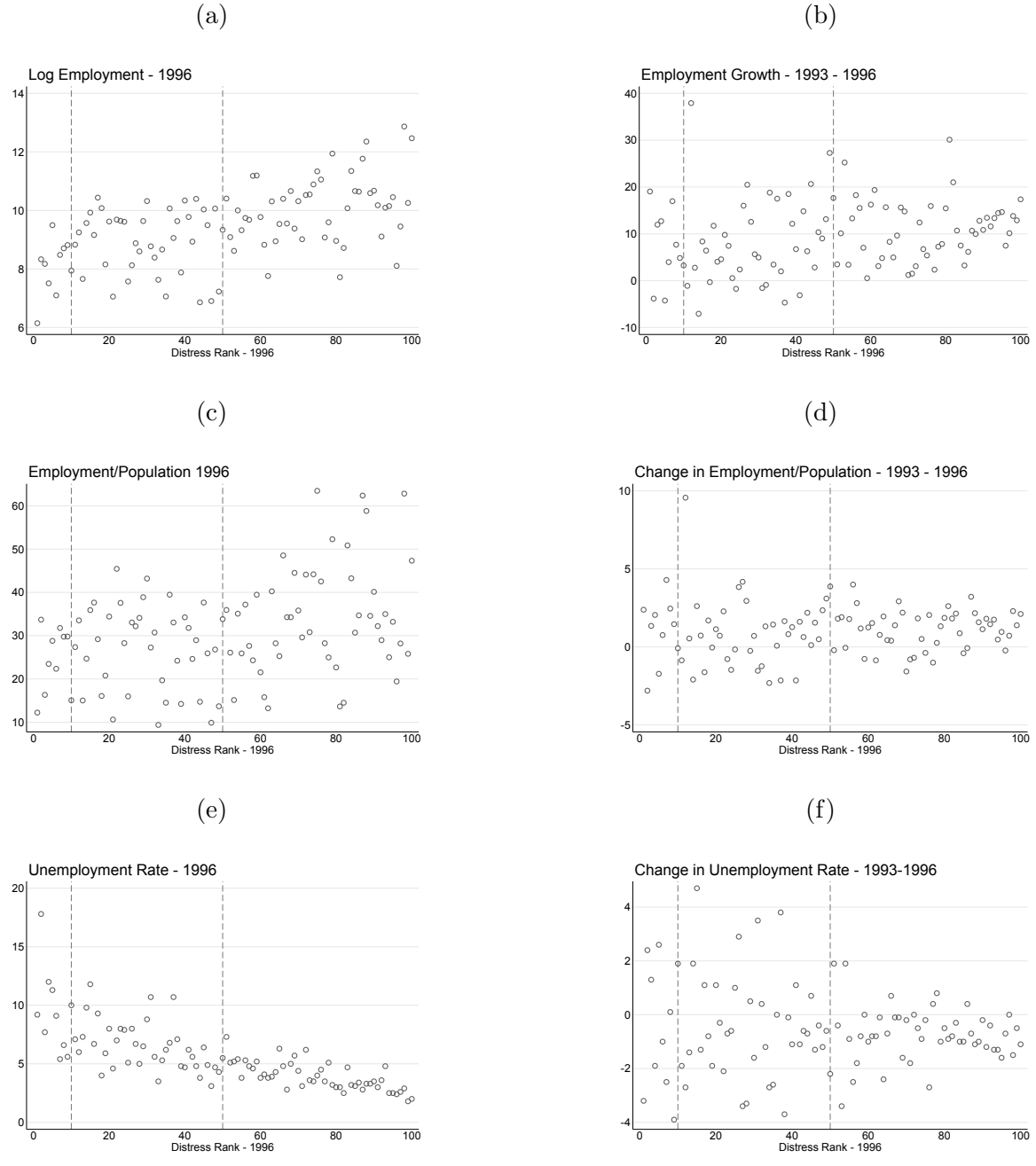
In this section, we discuss the difficulties of estimating the effect of hiring credits and lay out a strategy to take advantage of the assignment of subsidies based on distress ranks. North Carolina’s one hundred counties were assigned each year to three groups defined by the subsidy program tiers. 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 ten counties designated to Tier 1 in the initial program year, 1996, but distress rank ties and subsequent amendments to the program that added assignment rules lead this number to fluctuate between 10 and 28 in later program years. Tier 1 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 may apply for hiring credits ranging from 500 to 1 thousand dollars. We estimate the program’s effectiveness by comparing the evolution of employment and unemployment across counties in Tiers 1 and 2. To avoid making comparisons between significantly different counties, we exclude counties always designated as Tier 3. These contain major cities that 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 in them. However, this relationship is smooth across the tier cutoffs. Moreover, the slope of this relationship is small, suggesting that the correlation between the economic distress rank and these outcomes is not strong, both within each tier and across tiers for counties with similar distress ranks. The correlation between the distress rank and the change in the outcomes before program’s onset is also weak. However, the unemployment rate is more variable over time for more distressed, smaller counties.⁷

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 correlates weakly with economic variables. If counties that get assigned into Tier 1 have worse unobservables—that imply different trajectories of employment and unemploy-

⁷In Appendix Figure B.3, 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



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.

ment even in the absence of the program— then estimates from a difference-in-differences approach will be biased.

We address this possible bias by exploiting the discontinuities in tier assignments based on the economic distress rank. We follow [Cellini et al. \(2010\)](#) in implementing a dynamic regression discontinuity design. This estimation strategy exploits the yearly changes in tier assignments and allows for contemporaneous and lagged effects of the program.

Our baseline specification is:

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

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 the county, year, and time since treatment designation. $tier1_{c,t-k}$ is a dummy variable equal to 1 whenever a county is in 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 the controls, when included, are allowed to vary with time since treatment assignment k but do not vary by calendar year t . The variables $X_{c,t-K-1}$ are control variables measured at $K - 1$, before the program effects start.

Allowing for lagged effects is essential since the hiring subsidy programs may take a few years to gain traction and have a noticeable impact 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 program’s impact 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 program’s effect, 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. The estimation strategy addresses mean reversion bias even though Tier 1 assignment depends indirectly on pre-assignment outcomes 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 did not entirely depend on the economic distress rank. Counties could not be re-designated out of Tier 1 due to an improved distress rank until after two years. From 2000 onwards, high-poverty and low-population-based rules enter 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 that

are assigned based on additional rules and estimating equation (1) as a sharp discontinuity design using only counties assigned based on the running variable being considered, in this case, the distress rank. Another approach is to classify the counties that 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.⁸ 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 assuming constant treatment effects and a linear control function with the same slope on either side of the cutoff that is consistent with data before the beginning of the program.

If the program’s effect is constant over time, we can take advantage of its repeated execution. 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. we generate a new set of observations for outcomes stretching from two years before to four years after treatment designation for each program year and county. We pool together these spans of observations 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 assumption concerns 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. Therefore, we 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 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

⁸Tier 1 has counties ranked ten 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.

randomization inference (Cattaneo et al., 2015, 2016).

We also experiment with how we account for the dynamic nature of the subsidy program. Consider the effect of the program two years after its enactment in year t . In the 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 at $t + 1$. 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 that accounts for the indirect impact of initial treatment on the probability of treatment in future years.⁹ We apply their method with the following regression:

$$Y_{ct} = \beta_0 + \gamma_c + \gamma_t + \sum_{k=0}^K (\alpha_k m_{c,t-k} + \theta_k tier1_{c,t-k} m_{c,t-k} + \nu_k m_{c,t-k} f(rank_{c,t-k})) + X_{c,t-k-1} \beta + \varepsilon_{ct}. \quad (2)$$

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 $t - k$, so the county was either in Tier 1 or Tier 2 at the time. The θ_k coefficients measure the program impact at various lags. The coefficients on the ranking and 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 variation only 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. Including this history 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. We array counties by the initial distress rank relative to the threshold where credit

⁹The 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. There is full compliance with the hiring credits program at the county level when accounting for the overrides.

size increases. A linear fit in county rank is included, constrained to have the same slope on either side of the threshold. Given the prior evidence that program effects appear only with a lag, we focus on three-year differences. The figure pools outcomes for counties entering the program at different points in time. The outcomes correlate weakly with the distress rank, with the assumption of a linear relationship with rank appearing reasonable.¹⁰

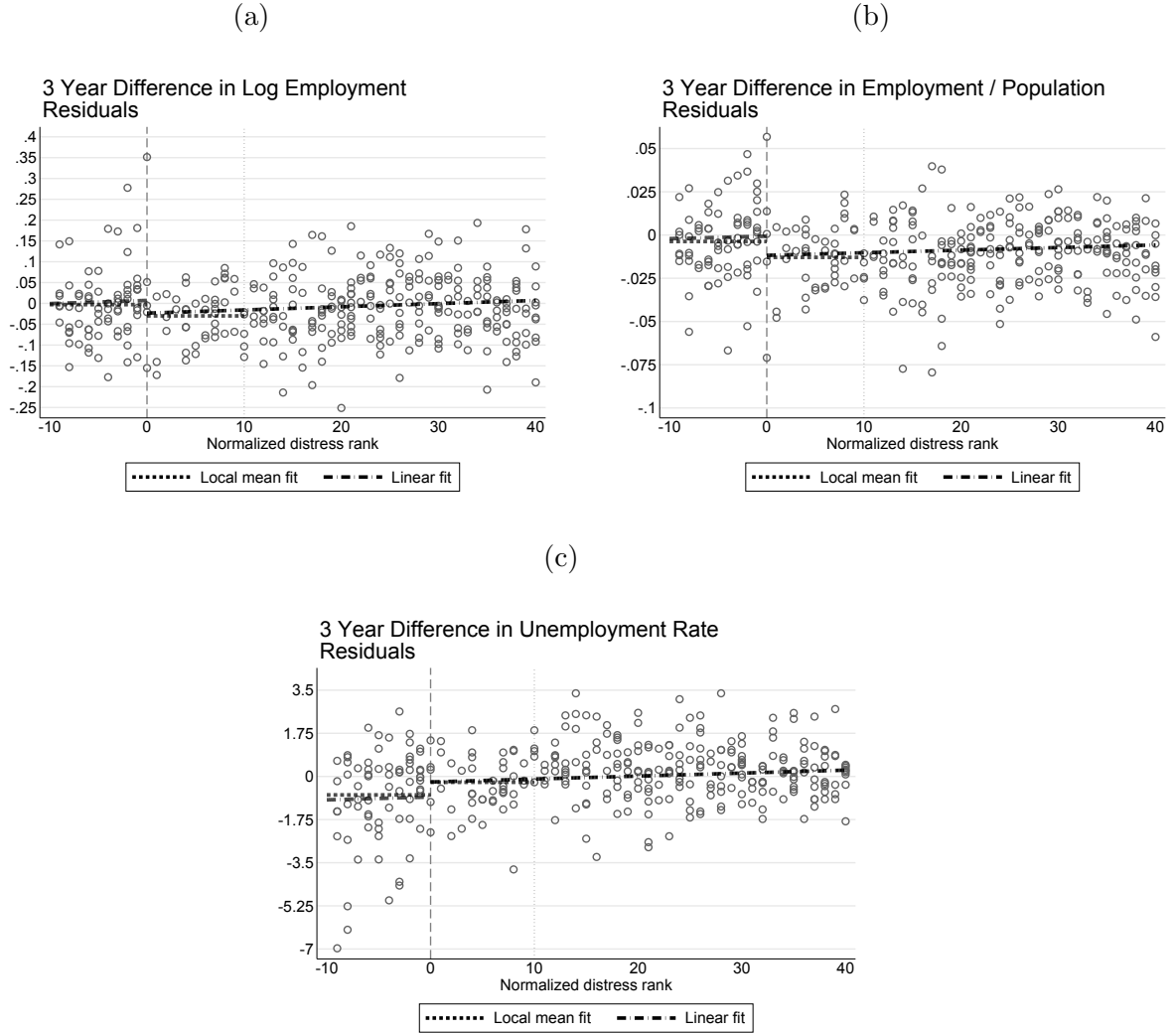
Table 2 presents the dynamic ITT estimates from equation (1). Rows 1 and 2 show estimates for log employment, which are progressively increasing for treated counties relative to the counterfactual through three years after treatment designation. The three-year effect with the ranking input variables as additional controls shows an employment increase of 3.6%. The same pattern of estimates emerges for the employment/population ratio in the next two rows. The effect three years after treatment is estimated to be around one percentage point. Rows 5 and 6 show that the unemployment rate in Tier 1 counties is smaller 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 being the result of adding additional county-level controls. For reference, throughout the program, unemployment rates averaged 6.6% for the sample overall and 7.9% for the most distressed counties comprising the treatment group.

The persistence of treatment status within counties further explains the presence of lagged effects. 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 the 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 it estimates extra parameters simultaneously.

¹⁰There 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 due 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 generated by Lowe’s in every county, but aggregate data indicate that Lowe’s generates 673 jobs in 2003, 271 in 2004, and 274 in 2005. This change in employment totals accounts for 1,218 direct hires generating credits, close to the entire employment increase shown in figure 5. In appendix table B.3 we show local estimates excluding this data point. The results are qualitatively similar.

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



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

Clustered standard errors in parentheses.

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.

On the other hand, the IV estimates show sizable effects three years after treatment, 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 than the employment results. Three years after, eligibility for the most extensive hiring credits has reduced the unemployment rate in a county by 0.5 percentage points according to the OLS specifications that exclude defiers. 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. The beneficial effects of initial treatment on unemployment rates should reduce the probability of receiving the program in future years. TOT estimates are larger for our employment estimates. For our unemployment rate estimates, the OLS ITT and TOT estimates are similar, but the IV TOT estimates are larger than both of them. This difference may be due to the lower efficiency and higher finite sample bias of IV estimates. The fact that the OLS TOT estimates are similar to the OLS ITT estimates implies that the program’s benefits 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 a random assignment of treatment at the policy threshold conditional on a control function that allows for more distant observations to contribute to the estimation of the treatment effect. As such, this estimation requires assumptions about the shape of that control function. Cattaneo et al. (2015) propose a non-parametric estimation technique that uses randomization inference in a small neighborhood around the threshold. We implement their approach in Table 4, where we show estimates for the program’s effect three years ahead.¹¹

To determine the estimation window where the random assignment assumption is deemed most plausible, we conduct a series of covariate balance tests for progressively larger bandwidths around the threshold. We use the lags of the income per capita, unemployment rate, share college, and target industry employment share as the covariates. Figure B.1 in the appendix visualizes the tests which yield a bandwidth of ± 6 . We then estimate the treat-

¹¹Full estimation results for one, two, and three years after treatment designation are in Table B.2 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 (0.016)	-0.005 (0.014)	0.009 (0.019)
Log Employment - IV	-0.065 (0.038)	0.038 (0.031)	0.072 (0.031)
Employment/Population - OLS	-0.008 (0.004)	-0.001 (0.004)	0.007 (0.005)
Employment/Population - IV	-0.016 (0.011)	0.010 (0.008)	0.023 (0.009)
Unemployment Rate - OLS	0.150 (0.343)	-0.582 (0.337)	-0.530 (0.306)
Unemployment Rate - IV	-0.154 (0.633)	-1.049 (0.498)	-1.144 (0.609)

Clustered standard errors in parentheses.

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.

ment effect as a simple difference in means, which we display as overlays to Figure 5. We calculate p-values 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 and windows of 10 and 20 ranks for more precision but 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, decreasing to about one 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 precise estimates of -0.5 to -0.9 percentage points, in line with the parametric

estimates. This change in the coefficients as the window expands implies that mean reversion in unemployment rates is upwardly biasing these estimates, particularly for the most severely distressed counties, ranked four and below, which are excluded from the narrowest bandwidth.

One concern with the local estimates, particularly for the smallest window, is the impact of introducing the population/poverty overrides to the ranking scheme that begins with the 2000 program year. Because population growth is an input to the distress rank, counties with the slowest growth are more likely to get a lower distress rank and 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. Once the overrides kick in after 2000, many of the lowest-performing counties to the right of the Tier 1 threshold leave the control group. This removal occurs because a lower distress rank correlates with the probability of being a defier. Such removal 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 earliest four cohorts from 1996-1999, whose tier designations are not subject to population/poverty overrides, observed three years later). Focusing on the six 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 to eligible industries, with the main sectors being manufacturing, warehousing, wholesale trade, and data processing.¹² 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 to population

¹²At 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.

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 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	0.067 (0.039)	0.030 (0.068)
Log Employment Non-Target Industries	0.003 (0.046)	-0.073 (0.090)
Employment/Population Target Industries	0.012 (0.004)	0.014 (0.008)
Employment/Population Non-Target Industries	-0.001 (0.009)	-0.016 (0.020)
Log Hires Annual Total	0.062 (0.040)	0.198 (0.091)
Log Separations Annual Total	0.043 (0.039)	0.178 (0.089)

Clustered standard errors in parentheses.

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 impact on the ratio of each industry’s employment to the overall county population. The three principal target industries are at the top. The effects are small and insignificant for all the sectors 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 dis-

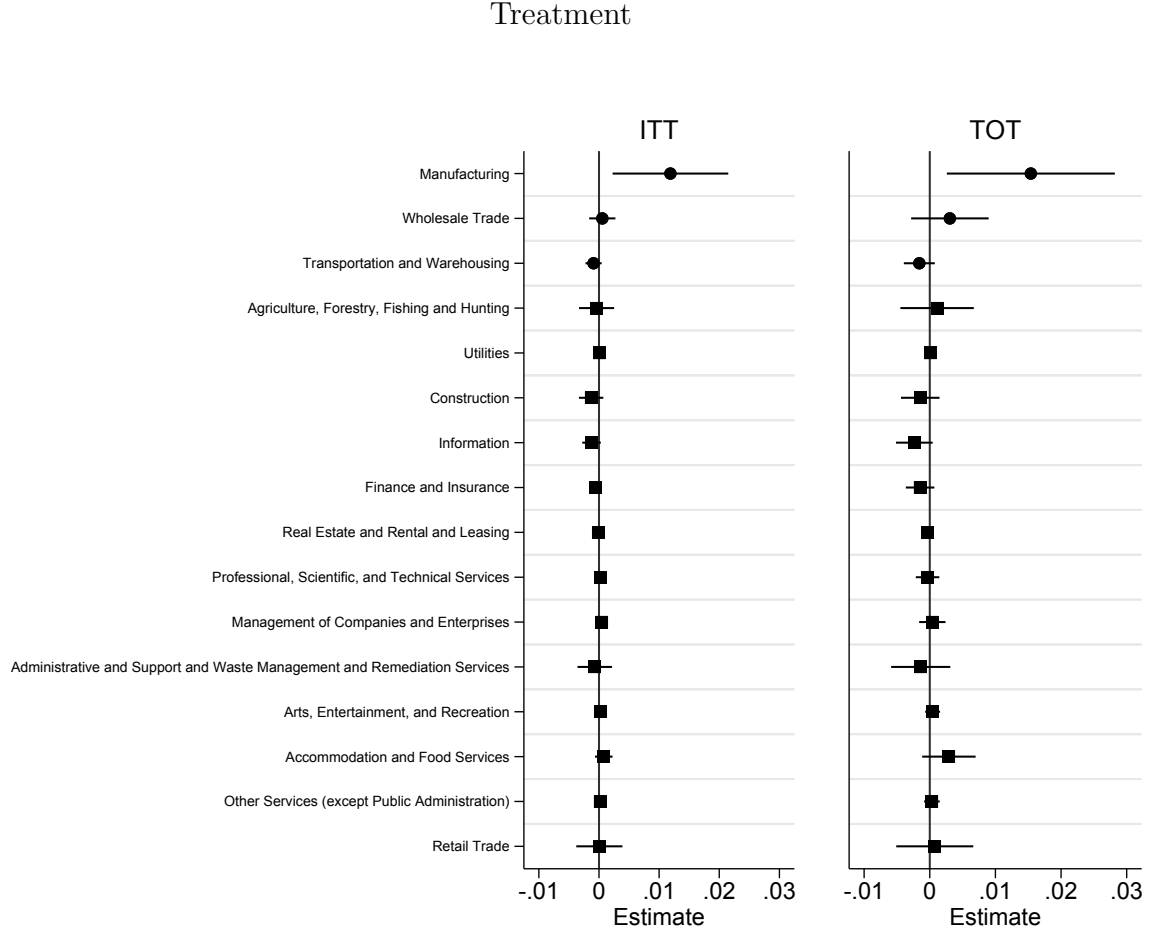
cernible impact for warehousing may be due to its combination with a non-targeted industry, transportation. With both wholesale trade and warehousing, another factor potentially limiting tax credit use was the wage requirement. Credits could only be claimed for jobs paying above the county average wage, 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 coefficients are consistent with employment gains resulting from increased hiring rather than decreased separations. Separations show evidence of an increase as well. This tracks with the empirical observation in the labor literature that growing firms continue to have positive separation rates (Cahuc et al., 2014).

5.3 Estimates for the Second Program Wave

Ideally, we would leverage Wave 2’s expanded eligibility for the largest hiring credits to quantify program impact heterogeneity as a function of initial economic distress. Unfortunately, the randomization at the eligibility threshold underpinning the RD design for Wave 1 estimates is not as supportable for Wave 2. This wave ran from 2007 to 2013, with a significant expansion of hiring credits. As shown in table 1, the number of counties eligible for the largest credit amounts increased from 10 to 40 counties. However, it is difficult to identify the effects of this second program wave because the impact of the first wave of the program induces differences in the evolution of labor market outcomes that change discontinuously across the distress rank cutoff of 40. We try to control for the differences caused by the first wave of the program, by estimating the effects of the second wave on a sub-sample of counties that did not have variation in credit size from 2003 to 2006. Nevertheless, there are substantial differences in covariates across the 40 distress rank cutoff in this sub-sample. Appendix figure B.2 shows a lack of balance in covariates for low bandwidths around the cutoff. This unbalance contrasts with the previous results for wave 1 in figure B.1. For completeness, in appendix table B.4 we show local estimates of the effects of the second wave of the program. These estimates show some significant reductions in unemployment, smaller than those found for the first wave.

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



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.

5.4 Difference-in-Differences Estimates

Our estimate of the employment impact of hiring credits of around 3% is in some contrast 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 targeted location. Besides the setting, prior studies have relied on difference-in-differences techniques that have the potential to be biased (positively or negatively) relative to a regression discontinuity approach. To assess the extent to which methodological differences underpin our larger impact estimates, we perform difference-in-differences (DD) estimation on the program.

Our basic specification is:

$$Y_{ct} = \beta_0 + \gamma_c + \gamma_t + \delta_c \times t + \sum_{k=0}^K \theta_k \text{tier}1_{t-k} + X_{c,t-K-1}\beta + \varepsilon_{ct}. \quad (3)$$

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 counties each year. $\text{tier}1_{t-k}$ 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.

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 that comprise the distress rankings' inputs. These controls address 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.1.

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 bias is consistent with the control function plotted in panel a of Figure 5 that 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

Table 6: Difference in Difference Estimates - Main Outcomes

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

Clustered standard errors in parentheses.

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.1 of the appendix.

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.

5.5 Program Costs

While these hiring subsidies have a positive impact on the labor market, a complete evaluation also needs to consider the policy’s cost. The cost accounting has some complications detailed below, but evidently, low wastage rates in this setting make the cost per job created look favorable. Lane and Jolley (2009) produced an evaluation of the incentive programs 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, the program generated \$2.1 billion in tax credits through its eleven-year run from 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 the taken credits could not exceed 50% of a firm’s annual tax liability, and unused credits could only be carried forward for five years. This paper’s focus, the job creation credits, comprised 17% of the total amount of generated credits.¹³ Though the program was more generous on a per-hire basis for the

¹³The 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

most distressed Tier 1 counties, firms in those counties received only 14% of generated credit dollars. Twenty-five percent of the credits went to Tier 2 counties, and over half of all credit dollars went to firms in the least distressed Tier 3 counties, given their greater size and hiring rates.¹⁴ 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.¹⁵

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 several factors in the design of the program that may have contributed to lower usage. 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 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. Sixty-two percent of surveyed executives at incented companies were unaware their company received an incentive. However, the number of firms in the least distressed tier where credits per hire were only \$500 was not reported (Lane and Jolley, 2009).

The issue of wastage, the share of expenditures for hires that would have taken place even in the absence of the policy change, 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 most benefits went to firms in economically robust counties where similar hiring and investment would likely have occurred absent the program.

To come up with a cost per job figure, we first generate the program impact on employment levels using the OLS log employment TOT specification from Table 3. We obtain the predicted increase in log employment for a county assigned to treatment three years prior but not in the intervening years (a 0.9% increase). Because the expenditure data is tier-wide, but the program impact is estimated only on non-defier counties, we need to assume a program benefit for that population. Because the defier Tier 1 counties tend to be less distressed, the

(1%).

¹⁴This breakdown is based on program years 2002 to 2006.

¹⁵We 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 for 2001 to 2006 multiplied by 0.65 for generated credits not taken, and divide by 11 for the per year figure.

program impact is likely smaller in those counties. We estimate that the differential generosity in credit size for Tier 1 relative to Tier 2 increased employment levels by 5,223 - 5,823 jobs three years later, corresponding to an assumed impact in defier counties ranging from zero to one as large as in non-defier counties. Adjusting the Tier 1 hiring credit expenditure data down to account for the total credit size (\$12,500) versus the differential credit generosity (\$9,000) yields a cost per job of around \$7,400 in nominal dollars or around \$12,000 in 2020 dollars. According to [Bartik \(2020\)](#), the benefit of a created job ranges from \$12,000 to \$48,000 making these hiring subsidies cost-effective in isolation, though potentially more costly than alternative job creation strategies like infrastructure investment.

Ideally, we would generate an estimate of wastage from this program by comparing our employment level estimate of around 5,500 jobs created to the 6,751 hiring credits claimed in Tier 1 throughout the program. This comparison would imply an atypically low windfall wastage rate of approximately 20%. The contrast is not clean, though, for two reasons pushing in opposite directions. First, the program’s impact theoretically combines direct and spillover job creation. That said, the results of [Table 5](#) and [Figure 6](#) indicate that spillover job creation was minimal. Second, because our program impact is for the differential credit generosity between tiers 1 and 2 rather than between a Tier 1 credit and nothing, the *total* impact of Tier 1 designation is ostensibly larger than the differential impact. Though we cannot arrive at a precise figure, it seems that a large majority of claimed credits were for hires induced by the program. Such low wastage contrasts with more broad-based employment subsidy programs where [Bartik \(2001\)](#) puts typical windfall wastage rates at around 70% and sometimes as high as 90%, drastically inflating the effective costs per job created. Moreover, if little of the credits claimed by firms in the more economically robust higher tiers induced job creation, the overall wastage rates for the hiring subsidies across the entire state would be closer to those cited figures.

5.6 Validity

While our use of an RD strategy should provide confidence in our estimates’ internal validity, some bias is still possible. Because subsidies are place-of-work based, 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 overestimating employment impacts and underestimating unemployment impacts.

There are many aspects specific to the program under study that may limit the results’ external validity. 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 effect 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.

Finally, the pairing of the hiring credits with other incentives for investment and R&D, while not atypical, means that 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 we focus on, we believe we are isolating a hiring credit effect.

6 Discussion

Hiring tax credits are a popular tool implemented at various times in many US states 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 prospects. We use 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 of the impact on employment are noisy but generally show a boost from the program of around 3%. In contrast, difference-in-differences estimates show no effect, in line with prior studies of hiring credits using cross-state comparisons. The downward bias in these latter estimates, relative to the more credible RD ones, is a reason to highlight the importance of accounting for time-varying unobservables. These unobservables may impact counterfactual county performance. [Cahuc et al. \(2019\)](#) finds consistent evidence of positive employment impacts of hiring credits at eligible firms, an impact estimate most analogous to our target industry results. Their estimates range from 0.8% to 2.3% for a credit about one fourth the size of what we study. Assuming the impact is linear in credit size, that finding is consistent with our target-industry employment growth estimates of around 7%.

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. The program’s windfall wastage rate appears to have been low, which may be due to the targeting of distressed areas where hiring in the absence of the program would have been limited. So even though inducing more economic activity in low productivity distressed areas may require

large incentives, the bulk of incentive expenditures are changing firms' behavior rather than being windfall payments. While hiring subsidies may not reverse regional divergence, they do have a significant positive impact. In principle, they can be designed in a cost-effective way if eligibility is restricted to the most distressed labor markets.

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.
- Bartik, Timothy J. (2020), "Smart Place-Based Policies Can Improve Local Labor Markets." *Journal of Policy Analysis and Management*, 39, 844–851.
- 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 (2019), "The Effectiveness of Hiring Credits." *Review of Economic Studies*, 86, 593–626.
- Cahuc, Pierre, Stéphane Carcillo, and André Zylberberg (2014), *Labor Economics, second edition*, volume 1. The MIT Press.
- 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 Rocio 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.
- Coate, Patrick and Kyle Mangum (2019), “Fast Locations and Slowing Labor Mobility.” Working Papers 19-49, Federal Reserve Bank of Philadelphia.
- 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.
- Kline, Patrick and Enrico Moretti (2014), “People, Places, and Public Policy: Some Simple Welfare Economics of Local Economic Development Programs.” *Annual Review of Economics*, 6, 629–662.

- 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 (2020), “Place-Based Policies: Can We Do Better Than Enterprise Zones?” *Journal of Policy Analysis and Management*, 39, 836–844.
- Neumark, David and Diego Grijalva (2015), “The Employment Effects of State Hiring Credits.” IZA Discussion Papers 9146, Institute for the Study of Labor (IZA).
- Neumark, David and Timothy Young (2019), “Enterprise zones, poverty, and labor market outcomes: Resolving conflicting evidence.” *Regional Science and Urban Economics*, 78.
- Program Evaluation Division (2015), “Final report to the joint legislative program evaluation oversight committee.” Technical Report 2015-11, North Carolina General Assembly.
- Saez, Emmanuel, Benjamin Schoefer, and David Seim (2019), “Payroll taxes, firm behavior, and rent sharing: Evidence from a young workers’ tax cut in sweden.” *American Economic Review*, 109, 1717–63.
- Sestito, Paolo and Eliana Viviano (2018), “Firing costs and firm hiring: evidence from an Italian reform.” *Economic Policy*, 33, 101–130.
- Slattery, Cailin (2020), “Bidding for Firms: Subsidy Competition in the U.S.” *Working Paper*.
- 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.
- Zabek, Mike (2019), “Local Ties in Spatial Equilibrium.” *Finance and Economics Discussion Series. Washington: Board of Governors of the Federal Reserve System*, 2019-080.

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 (1)	rank (2)	per cap (3)	rank (4)	rate (5)	rank (6)	sum (7)	rank (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

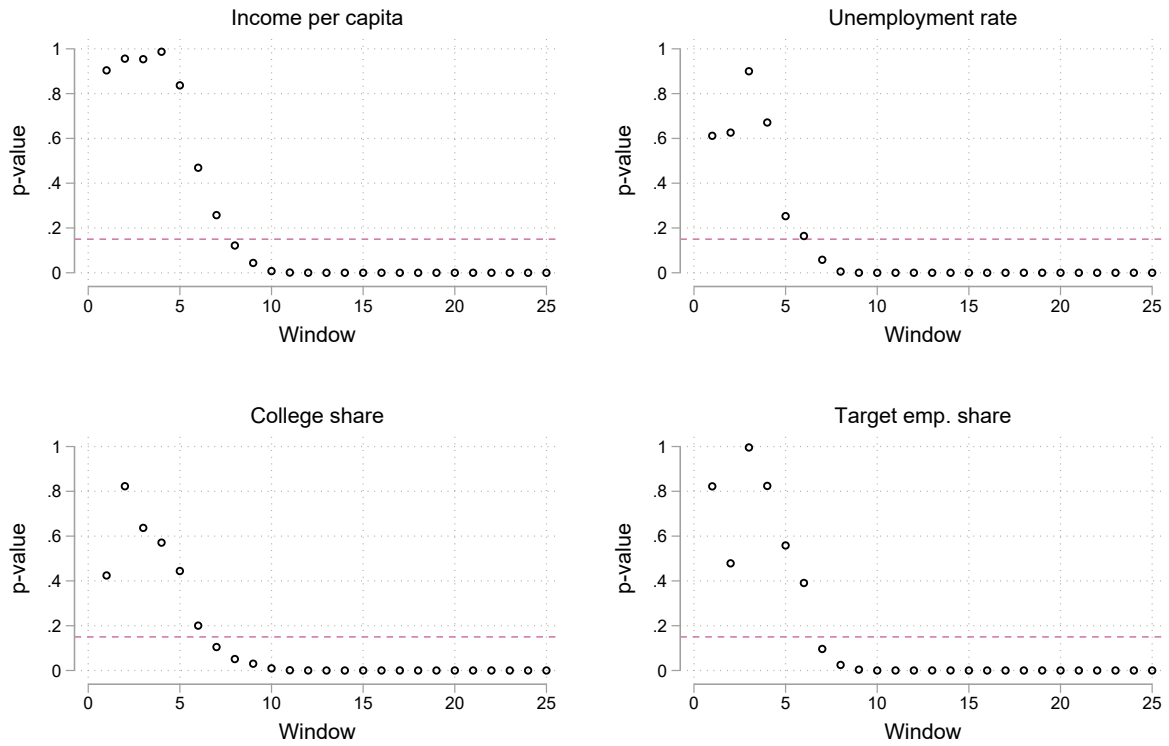
Table B.1: Difference-in-Differences Estimates

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

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.

Figure B.1: Local Estimates Window Selection



Note: Covariate balance tests by window size against the 15% level threshold recommended by [Cattaneo et al. \(2015\)](#).

Table B.2: 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 [0.906]	0.000 [0.874]	-0.198 [0.393]	88
	2 Years	0.026 [0.157]	0.007 [0.126]	-0.287 [0.392]	78
	3 Years	0.053 [0.018]	0.014 [0.013]	-0.050 [0.884]	69
10 ranks	1 Year	-0.006 [0.501]	-0.000 [0.949]	-0.135 [0.436]	152
	2 Years	0.010 [0.438]	0.004 [0.179]	-0.495 [0.066]	135
	3 Years	0.028 [0.089]	0.010 [0.022]	-0.508 [0.128]	119
20 ranks	1 Year	-0.004 [0.530]	0.000 [0.757]	-0.245 [0.046]	236
	2 Years	0.006 [0.556]	0.004 [0.088]	-0.643 [0.000]	211
	3 Years	0.025 [0.042]	0.010 [0.002]	-0.908 [0.000]	186

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

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

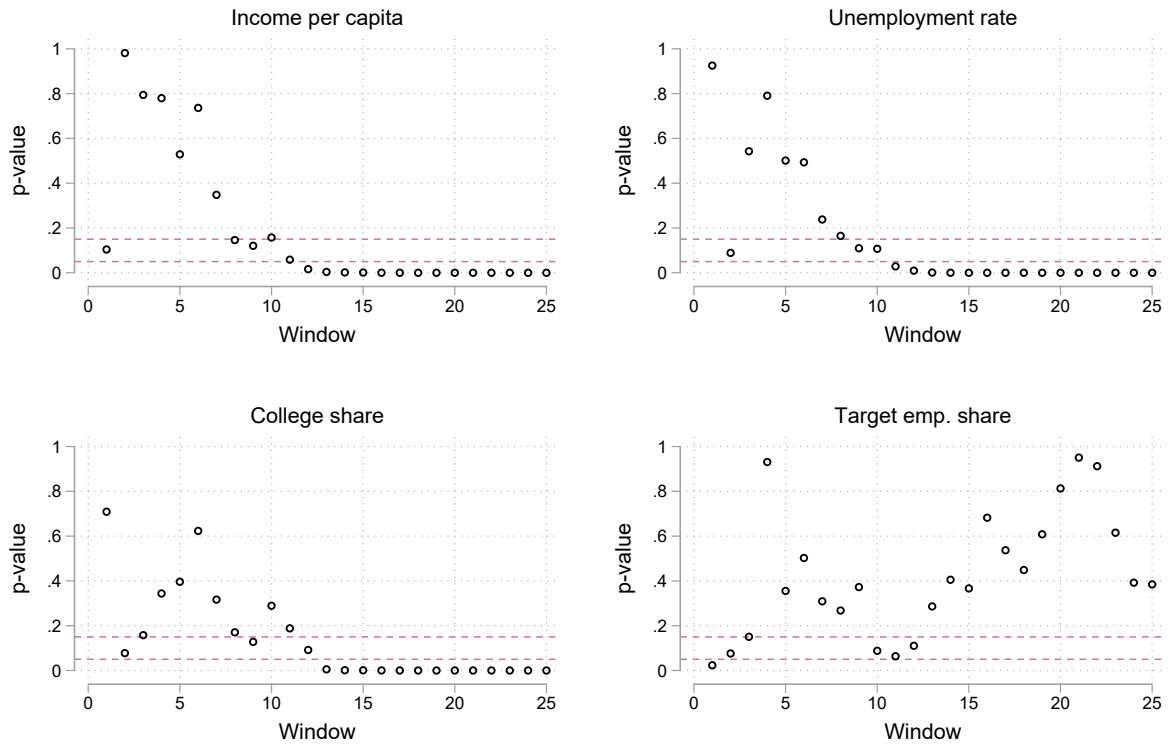
Table B.3: Local Estimates of Effect 3 Years after Treatment, Excluding Northampton
County in 2005

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.

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.

Figure B.2: Local Estimates Window Selection. Second Program Wave.



Note: Covariate balance tests by window size against the 15% level threshold recommended by [Cattaneo et al. \(2015\)](#).

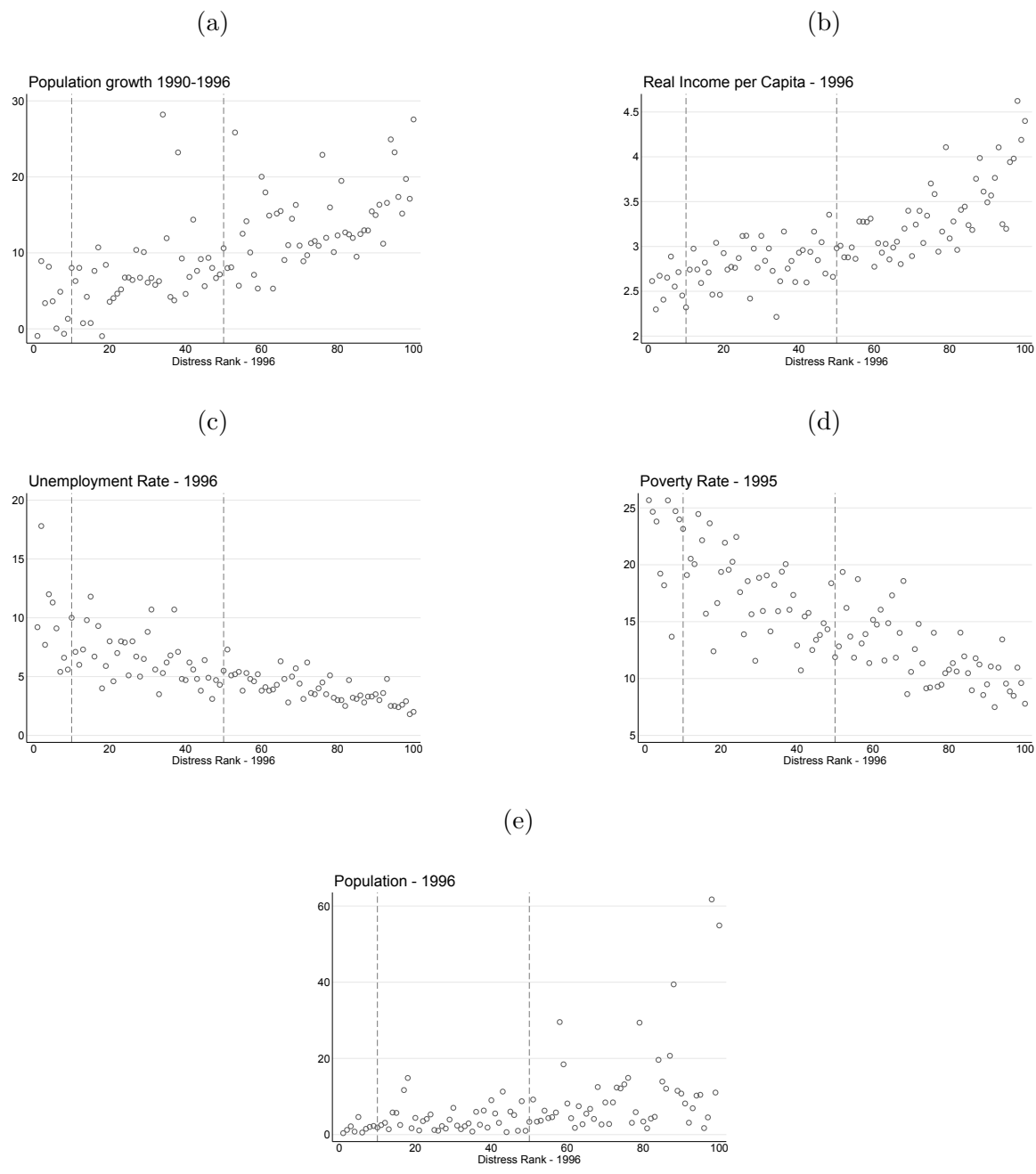
Table B.4: Local Estimates, Second Program Wave: 1,2 and 3 Years after Treatment

Window	Years	Dependent Variable			N
		Log Employment	Employment/ Population	Unemployment Rate	
6 ranks	1 Year	0.005 [0.628]	0.001 [0.644]	-0.409 [0.076]	42
	2 Years	0.000 [0.967]	0.000 [0.925]	-0.602 [0.066]	42
	3 Years	0.005 [0.750]	0.001 [0.824]	-0.712 [0.035]	42
10 ranks	1 Year	-0.008 [0.272]	-0.002 [0.325]	-0.241 [0.122]	73
	2 Years	-0.005 [0.689]	-0.000 [0.912]	-0.451 [0.057]	73
	3 Years	-0.005 [0.739]	-0.000 [0.945]	-0.671 [0.006]	73
20 ranks	1 Year	0.003 [0.641]	0.001 [0.562]	-0.315 [0.002]	156
	2 Years	0.003 [0.727]	0.002 [0.481]	-0.371 [0.026]	156
	3 Years	0.005 [0.588]	0.003 [0.267]	-0.590 [0.001]	156

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

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

Figure B.3: Relationship between the Distress Rank and Ranking Inputs



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.