

The efficacy of hiring credits in distressed areas

Jorge Pérez*and Michael Suher†

September 2018

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. We estimate the impact of the credits using difference in differences and regression discontinuity methods. Our estimates show fairly sizable and robust impacts on unemployment - a \$9,000 credit leads to a nearly 1 percentage point reduction in the unemployment rate. The attendant increase in employment levels is less precisely estimated but appears to be around 2%.

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

†Federal Reserve Board (e-mail: michael.suher@frb.gov). The views expressed are those of the authors and not necessarily those of the Federal Reserve Board or Banco de México.

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 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 include thresholds at which credit size jumps discontinuously allowing for difference in differences (DD) as well as regression discontinuity (RD) estimates. Our DD estimates show zero or occasional negative impacts on employment. The RD estimates are noisy but show employment increases of around 2%, indicating the importance of accounting for time varying unobservables. We find significant reductions in unemployment rates which grow over time, and are on the order of 1 percentage point for a \$9,000 credit.

A few 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 (2015) 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.

Bartik (2001) notes that when made generally available, a large share of credits ends up as wastage, meaning they are claimed by firms whose hiring would have taken place absent the program. 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 simultane-

ously increasing separations. Because our analysis focuses on overall employment levels and unemployment rates, estimates of program effects will be net of any churning. 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.

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 decadal frequency, they may fail to capture shorter run counter-factual employment trends, particularly during recessions. Our use of a regression discontinuity design can arguably offer more causal estimates of program effects. We also extend the analysis by considering the impact of hiring credits on unemployment rates.

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.

Chirinko and Wilson (2016) also 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.

Busso et al. (2013) and Freedman (2013) both assess the impact of the federal enterprise zone program which includes hiring credits, and find positive employment effects for neighborhood residents. The policy we investigate differs in that it targets a broader county rather than a specific neighborhood.

Our study builds on this prior research on the efficacy of hiring tax credits by incorporating plausibly causal regression discontinuity estimates and considering the size of credits

made available. We also look at unemployment rates in addition to the conventional focus on employment growth.

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 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. In its place, a revamped program was launched in 1996 known as Article 3A or the William S. Lee program. 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. This latter program operated in similar fashion to its predecessors, but with some changes to the credit eligibility formulas. Tier 1 - the most distressed - was 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 also amended to incorporate property value per capita alongside unemployment rates, income, and population growth. In 2014, the Article 3J program ended and was not replaced (Program Evaluation Division,

2015). The tax credit size histories are summarized in table 1.¹

Table 1: Credit size by distress rank (Dollars)

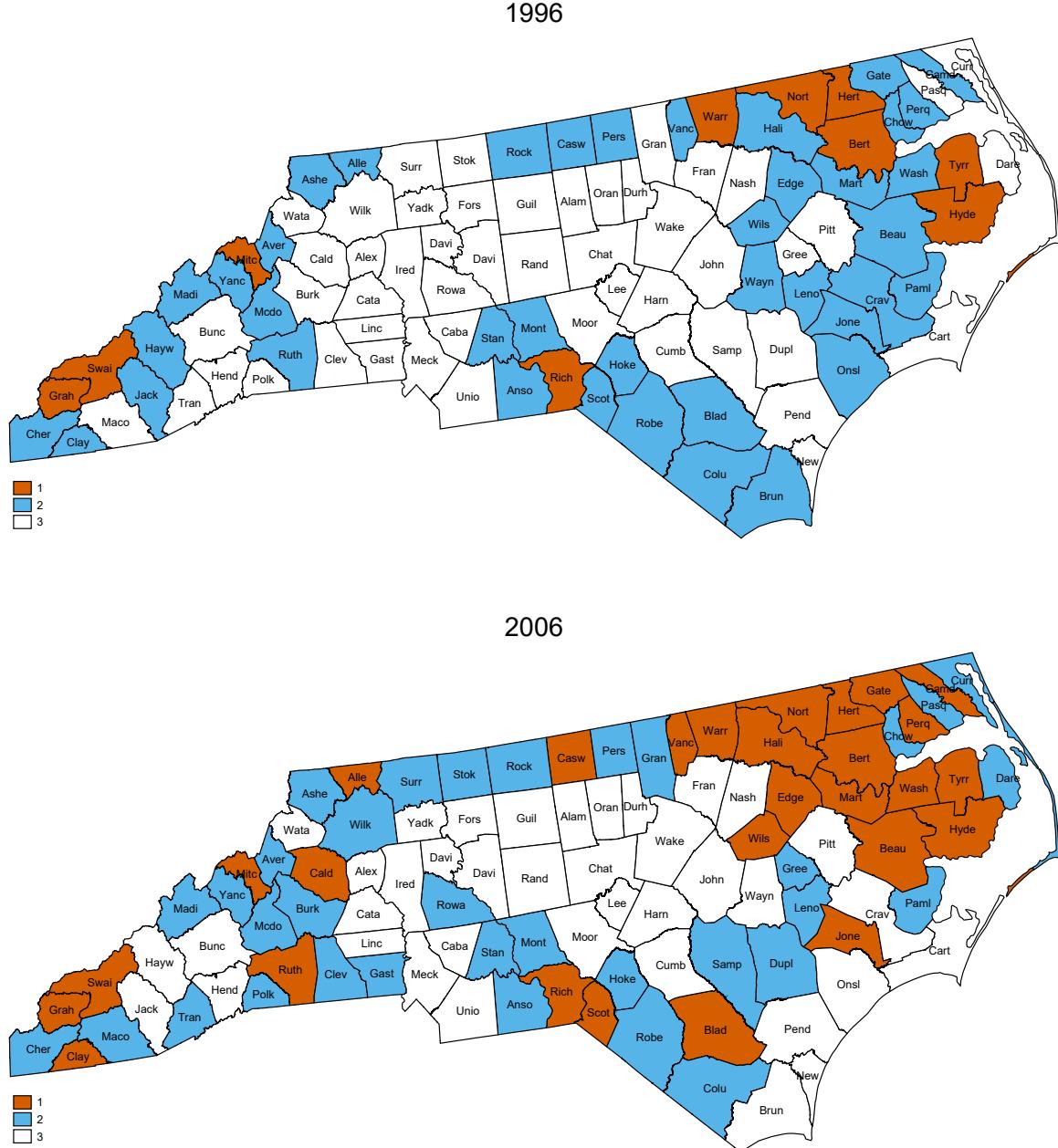
Years	Distress									
	10	20	30	40	50	60	70	80	90	100
wave0: 1988-1995	2,800									
wave1: 1996-2006	12,500	3,000-4,000				500-1,000				
wave2: 2007-2013	12,500			5,000				750		

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. 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. While there was no requirement that hires be of certain types of workers, such as those currently unemployed, it was restricted based on industry. The main industries eligible were manufacturing, wholesale trade, warehousing, and those related to data processing.

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. That is, the two counties ranked 10 and 30 can be quite different from each other, while the two counties ranked 70 and 90 are fairly similar to each other, or vice versa. A more continuous and robust measure of distress was proposed by the state Department of Commerce though not adopted (Department of Commerce, 2014). Below 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.

¹Under 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.

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.

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 2 shows overall conditions in North Carolina's economy during the program period. Population is steadily growing throughout. Employment is growing through 2000. It fluctuates through the 2000s, and returns to about the level of 2000 by the end of the period.

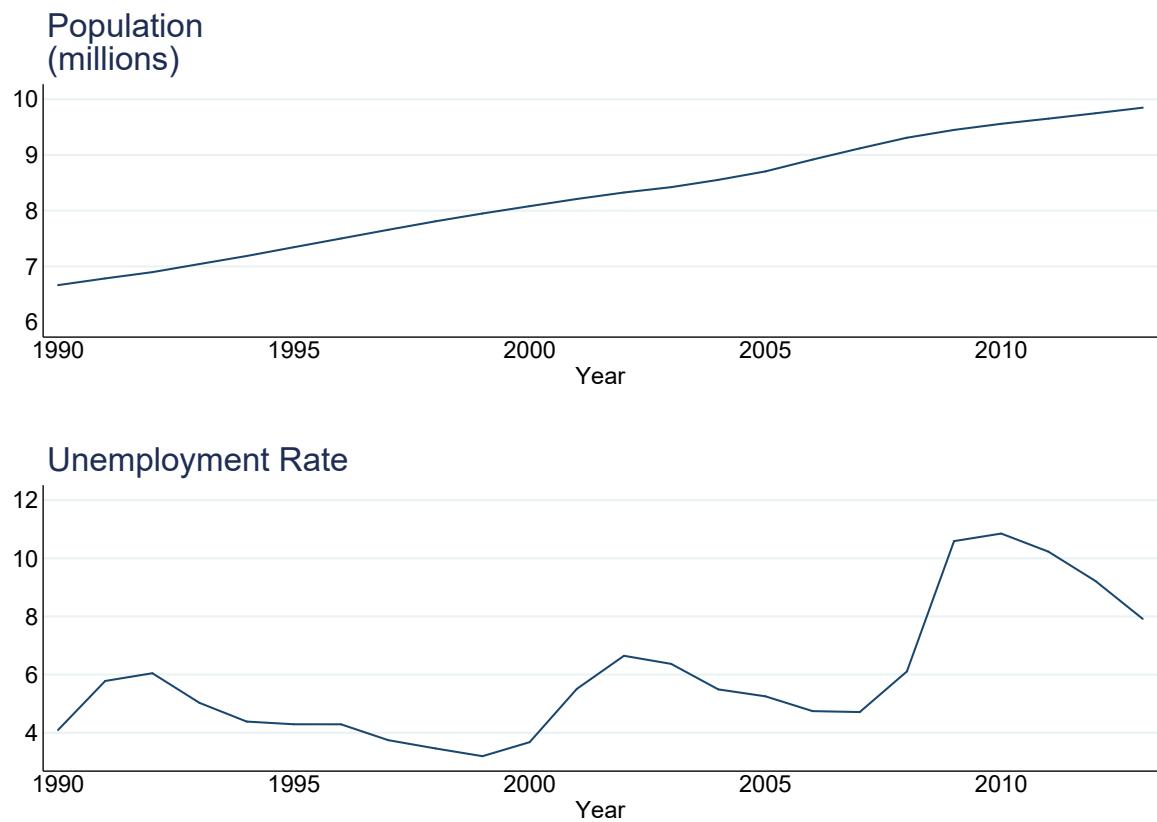
4 Estimating the effect of hiring credits

In this section we describe our estimation strategy. We lay out the difficulties of estimating the effect of hiring credits and suggest two strategies that take advantage of the assignment of subsidies based on distress ranks. First, we outline a difference in differences estimation strategy that compares counties across different tiers. Then, we describe a regression discontinuity design strategy that exploits changes in subsidy status across distress ranking thresholds.

4.1 Difference in differences estimates

North Carolina's 100 counties are 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

Figure 2: 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.

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 3 shows the relationship between several variables and the economic distress rank. Two messages emerge from these figures. As expected, there is an overall negative relationship between economic outcomes and economic distress. Poverty rates are higher for more distressed counties, while income per capita and population growth are 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.

Our basic specification is:

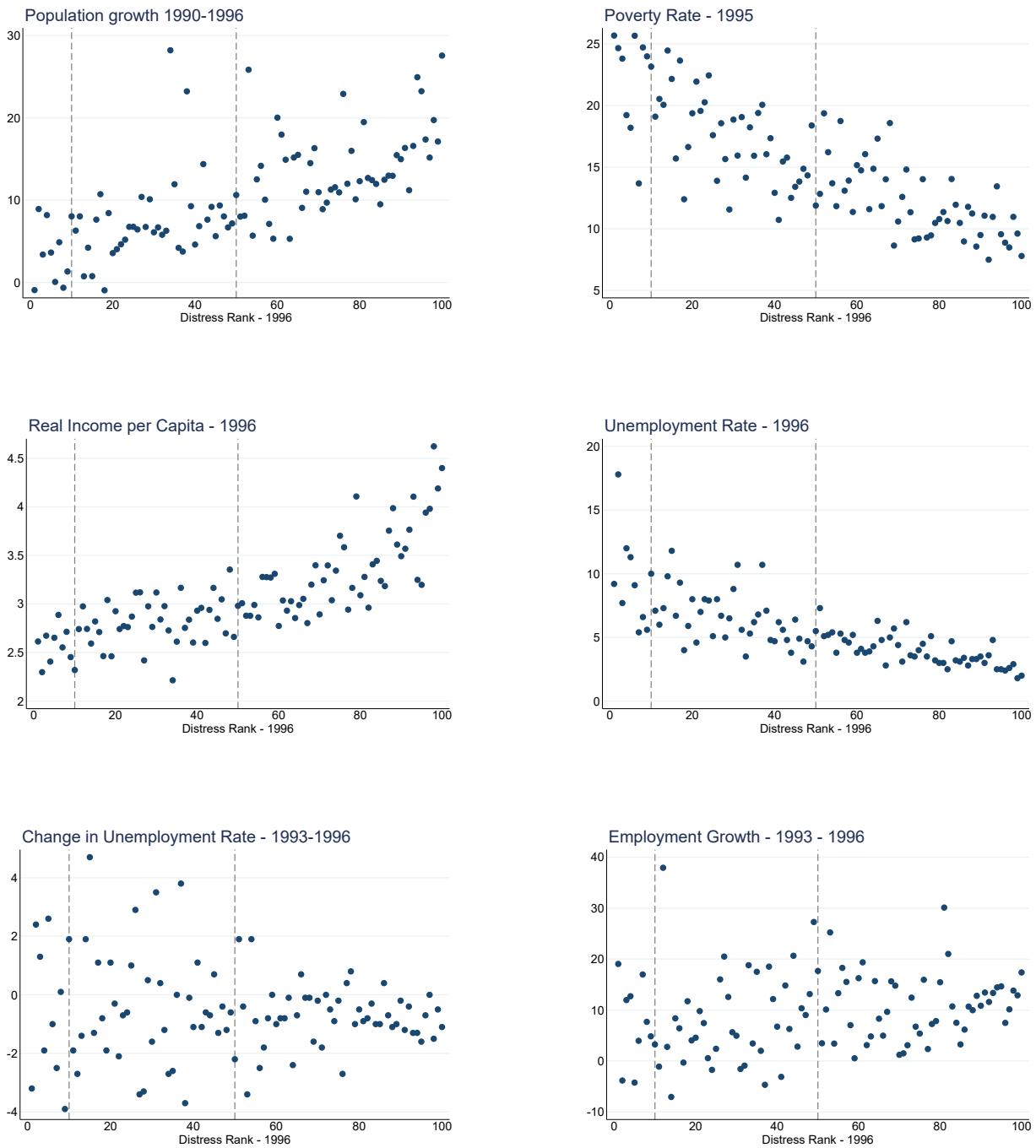
$$Y_{ct} = \beta_0 + \gamma_c + \gamma_t + \sum_{k=0}^K \theta_k tier1_{t-k} + \beta X_{ct} + \varepsilon_{ct} \quad (1)$$

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. 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.

This difference in differences strategy is valid only if unobservables have a similar evolution over time across tiers. We validate whether this is the case by looking at the evolution of the outcomes across different 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 tier and county-specific linear time-trends.

We calculate clustered standard errors by county to allow for serial correlation in the error term ε_{ct} within counties (Bertrand et al., 2004). Some of our regressions have less than

Figure 3: Relationship between distress rank, rank inputs 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.

50 counties. In those settings, clustered standard errors may be biased downward. Therefore, we also calculate p-values using a wild cluster bootstrap (Cameron and Miller, 2015) and report clustered standard errors and significance tests from the bootstrap p-values.

4.2 Regression discontinuity estimates

The main issue with the DD strategy outlined above is the possibility of correlation between ε_{ct} and the tier 1 variables. 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 of employment and unemployment even in absence of the program, then our estimates 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.

Our baseline specification is:

$$Y_{ctk} = \beta_0 + \gamma_c + \gamma_t + \gamma_k + \theta_k tier1_{c,t-k} + \nu_k f(rank_{c,t-k}) + \beta_k X_{c,t-k} + \varepsilon_{ctk} \quad (2)$$

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. The coefficients on the ranking and on the controls when included are allowed to vary with time since treatment assignment k .

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 (2) 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.

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.² 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 (as assumed in the DD estimates), 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 3 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).

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

²Tier 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.

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. We apply their method with the following regression:

$$Y_{ct} = \beta_0 + \gamma_c + \gamma_t + \sum_{k=0}^K (\alpha_k m_{c,t-k} + \theta_k tier1_{c,t-k} m_{c,t-k} + \nu_k m_{c,t-k} f(rank_{c,t-k}) + \beta X_{c,t-k}) + \varepsilon_{ct} \quad (3)$$

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 treatment on the treated estimate (TOT). This is meant to isolate the impact of having received treatment k years ago and not in subsequent years.

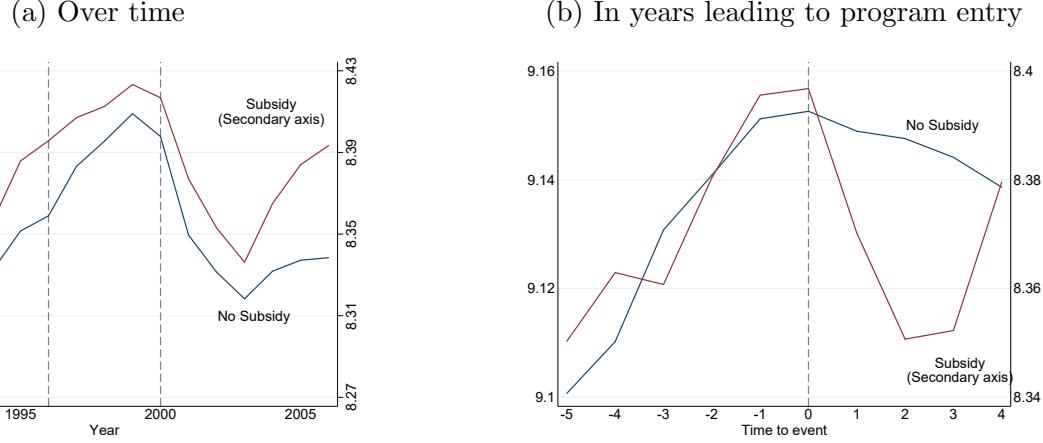
5 Results

5.1 Difference in differences estimates

Figure 4 portrays the DD analysis for log employment. As can be seen in panel (a), prior to the inception of the program in 1996, subsidy counties - those set to become eligible for the largest credits, had employment growth which was somewhat faster than the control counties, with the control counties catching up afterwards. Since the number of counties eligible for treatment is increasing over time, panel (b) looks at log employment centered on the year of entry to the program. The no subsidy line representing control counties is built by pooling together a different set of control counties for each possible year of treatment entry, weighted by the number of counties entering that year. Although in panel (a) there

seems to be a difference in trends between never treated counties and control counties, this trend is not apparent when we compare counties in years leading to program entry. We address some of this trend heterogeneity including lagged dependent variables and county level trends in our specifications. The later RD analysis is designed to be robust to this issue without the need for further assumptions or controls.

Figure 4: Log employment in treated and control counties



Note: Data is from the Bureau of Labor Statistics. Firms in subsidy counties are eligible for the largest size hiring tax credits. Vertical lines denote the initiation of the program in 1996 and the significant expansion in counties eligible for the program in 2000. Left and right axes are adjusted to the same scale.

Figure 5 portrays the DD analysis for unemployment rates. Here pre-trends are more parallel making the difference in differences estimates more credible. A program effect is also apparent here with subsidy county unemployment rates falling after entry into the program, though with a lag.

Table 2 implements the DD strategy for log employment. Column 2 introduces lags of treatment status. Column 3 allows for heterogeneous trends at the treatment/control group level. Column 4 allows each county to have its own trend. Column 5 includes lagged time varying county level controls. Even once these are added, estimated program impacts are still zero or even negative. Theoretically, the program impact could be zero but not negative. We interpret these estimates as indicating the presence of heterogeneous trends in employment levels for which the differencing strategy cannot account.

Table 3 presents the DD estimates for unemployment. Focusing on column 5, the estimated impact of the program occurs with a lag. Cumulatively, in the fourth year of exposure to the program, unemployment rates are lowered by about 0.7 percentage points. For ref-

Table 2: Difference in difference estimates: Employment

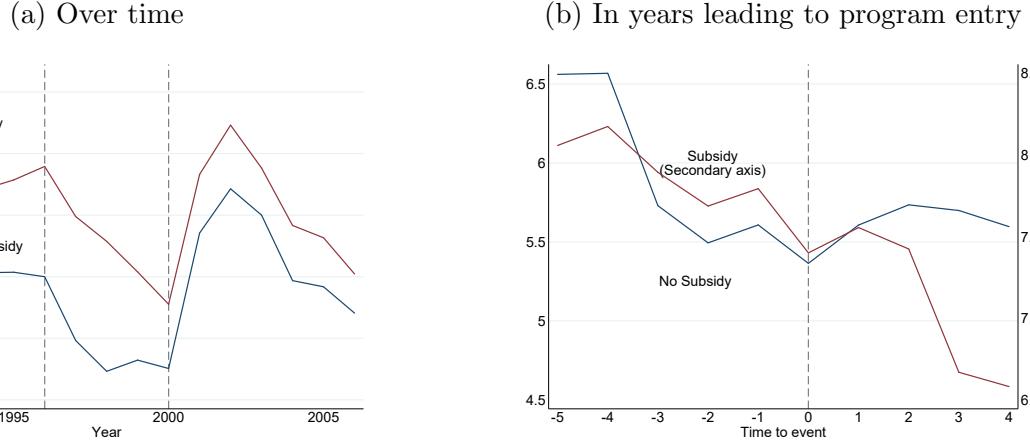
	Dependent Variable: Log Employment				
	(1)	(2)	(3)	(4)	(5)
Tier1	-0.00398 (0.0186)	0.00699 (0.0136)	0.00356 (0.0118)	-0.00652 (0.00870)	0.000638 (0.00825)
Lag Tier 1		-0.0225 (0.0131)	-0.0239* (0.0138)	-0.0206** (0.00904)	-0.0162* (0.00905)
Lag 2 Tier 1		-0.00488 (0.00750)	-0.00616 (0.00768)	-0.0143 (0.00898)	-0.0107 (0.00849)
Lag 3 Tier 1		0.0208 (0.0164)	0.0156 (0.0172)	-0.00152 (0.0139)	0.000525 (0.0133)
Lag 4 Population growth					-0.0000861 (0.00303)
Lag 4 Real Income per capita					0.0199 (0.0453)
Lag 4 Unemployment Rate					-0.00560* (0.00307)
<i>R</i> ²	0.991	0.994	1.000	1.000	1.000
N	714	588	588	588	546
Counties	42	42	42	42	42
County_FE	Yes	Yes	Yes	Yes	Yes
Year_FE	Yes	Yes	Yes	Yes	Yes
Group_trends			Yes		
County_trends				Yes	Yes

Clustered standard errors in parentheses

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

Note: Table depicts difference in difference estimates of equation 1. 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. Column (3) includes a linear time trend interacted with the Tier 1 dummy. Columns (4) and (5) include linear time trends interacted with county dummies.

Figure 5: Unemployment rate



Note: Data is from the Bureau of Labor Statistics. Firms in subsidy counties are eligible for the largest size hiring tax credits. Vertical lines denote the initiation of the program in 1996 and the significant expansion in counties eligible for the program in 2000. Left and right axes are adjusted to the same scale.

erence, 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.

5.2 Regression discontinuity estimates

We now turn to the regression discontinuity estimates. Figure 6 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 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.

Table 4 presents the dynamic ITT estimates from equation (2) for log employment in rows 1 and 2 and unemployment in rows 3 and 4. Employment is 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 2.6% though this is not statistically significant. The unemployment rate in tier 1 counties is reduced relative to control counties even one year after treatment and continues its relative decline through three years after. Relative to control counties, the

Table 3: Difference in difference estimates: Unemployment

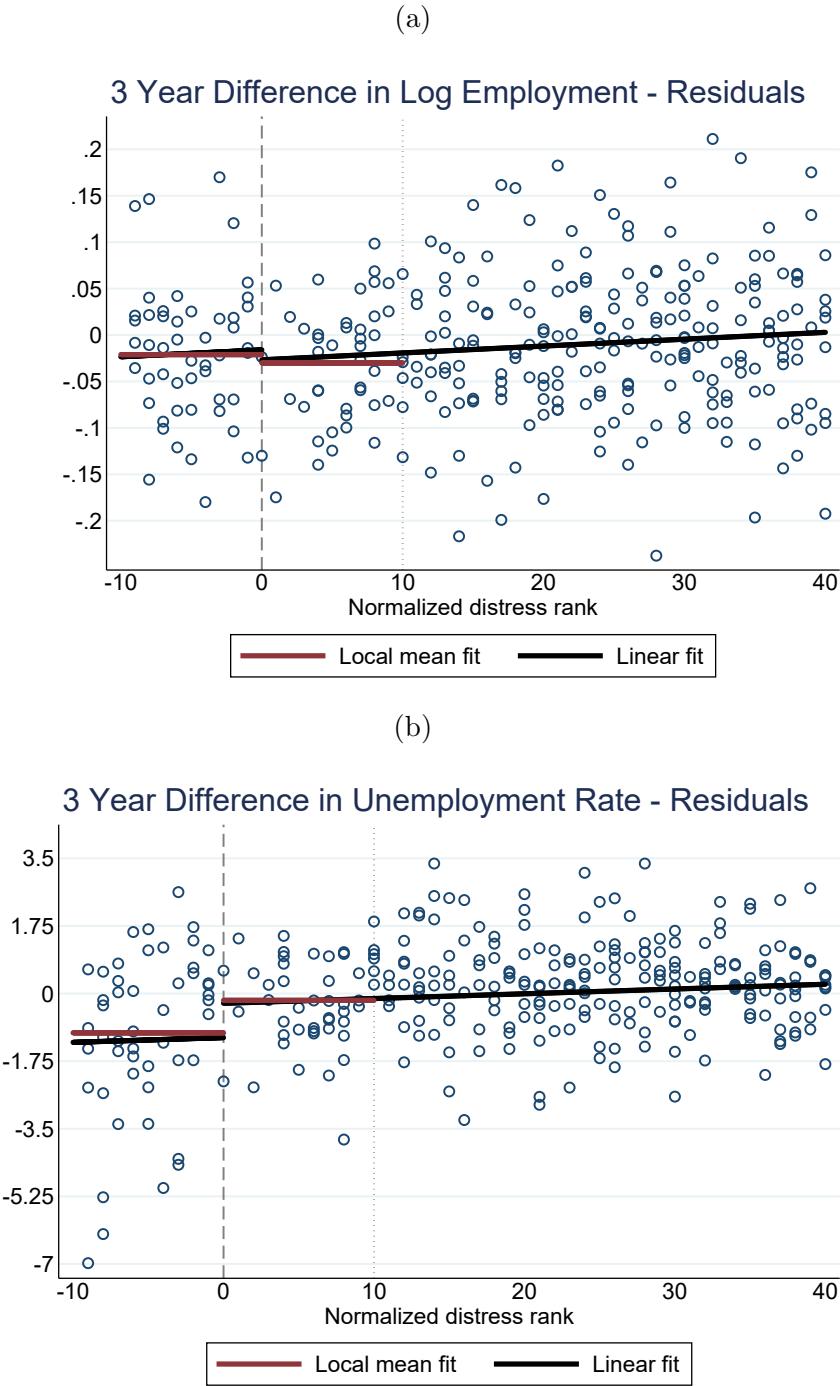
	Dependent Variable: Unemployment Rate				
	(1)	(2)	(3)	(4)	(5)
Tier1	-0.0532 (0.283)	0.264*** (0.187)	0.273 (0.187)	0.0724 (0.164)	0.0774 (0.159)
Lag Tier 1		-0.0793 (0.200)	-0.0758 (0.204)	-0.121 (0.168)	-0.0512 (0.186)
Lag 2 Tier 1			-0.383 (0.216)	-0.380* (0.207)	-0.293*** (0.117)
Lag 3 Tier 1				-0.956* (0.314)	-0.943*** (0.313)
Lag 4 Population growth					0.0415 (0.0303)
Lag 4 Real Income per capita					0.655 (0.694)
Lag 4 Unemployment Rate					-0.111** (0.0463)
<i>R</i> ²	0.652	0.669	0.969	0.982	0.984
<i>N</i>	714	588	588	588	546
<i>Counties</i>	42	42	42	42	42
County_FE	Yes	Yes	Yes	Yes	Yes
Year_FE	Yes	Yes	Yes	Yes	Yes
Group_trends			Yes		
County_trends				Yes	Yes

Clustered standard errors in parentheses

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

Note: Table depicts difference in difference estimates of equation 1. 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. Column (3) includes a linear time trend interacted with the Tier 1 dummy. Columns (4) and (5) include linear time trends interacted with county dummies.

Figure 6: Discontinuities in 3 year differences of employment and unemployment.



Note: Scatterplots of 3-year differences in outcomes, 1996-2006. Figure plots the sample mean plus the 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.

unemployment rate in treated counties is between 0.9 and 1.3 percentage points lower after three years with the lower estimate the result of adding additional county level controls.

Table 4: Regression discontinuity ITT estimates - Main outcomes

	1 yr later	2yrs later	3 yrs later
Log Employment	0.008 (0.013)	0.011 (0.015)	0.024 (0.020)
with controls	0.006 (0.013)	0.011 (0.015)	0.026 (0.020)
Unemployment Rate	-0.531* (0.270)	-1.050*** (0.323)	-1.324*** (0.391)
with controls	-0.353 (0.262)	-0.750*** (0.230)	-0.888*** (0.213)

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 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 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.

Tables 5 and 6 implement equation (3) for the dynamic TOT estimates using the two different approaches proposed by Wong et al. (2013) for handling multiple assignment rules. Column 1 excludes from treatment counties assigned by rules besides the distress rank. Column 3 includes all counties assigned to tier 1 by any rule, and instruments for treatment using the primary assignment rule based on distress rank. Respective estimations including county level controls in addition to the running variable are reported in columns 2 and 4. For employment growth, the results of treatment are generally statistically insignificant. This likely reflects the more demanding nature of this estimation relative to the ITT specification above as more parameters are being estimated simultaneously. In table 6 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 1.2 to 1.3 percentage points, though are not statistically significant.

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. The fact that the TOT estimates are largely similar to the 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 don't 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.

Table 5: Regression discontinuity TOT estimates: Employment

	Dependent Variable: Log Employment			
	(1) OLS	(2) OLS	(3) IV	(4) IV
Lag 1 Tier 1	-0.019 (0.017)	-0.019 (0.021)	-0.069* (0.039)	-0.075 (0.065)
Lag 2 Tier 1	-0.025** (0.012)	-0.019 (0.016)	0.029 (0.035)	0.039 (0.049)
Lag 3 Tier 1	-0.013 (0.019)	-0.012 (0.022)	0.006 (0.055)	0.003 (0.079)
<i>R</i> ²	0.997	0.997	0.187	0.173
<i>N</i>	770	700	770	700
<i>Counties</i>	70	70	70	70
<i>Controls</i>		Yes		Yes

Clustered standard errors in parentheses

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

Note: Table depicts estimates of equation 3. Standard errors are clustered by county. All columns include time and county fixed effects. Columns (2) and (4) include lags of the unemployment rate, income per capita, and population growth as additional controls. Columns (1) and (2) exclude from estimation of the treatment effects any counties designated as Tier 1 by an assignment rule besides the primary one based on distressed rank. Columns (3) and (4) label all Tier 1 counties as being treated and instrument for treatment with the distressed rank assignment rule.

5.3 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

Table 6: Regression discontinuity TOT estimates: Unemployment

	Dependent Variable: Unemployment Rate			
	(1) OLS	(2) OLS	(3) IV	(4) IV
Lag 1 Tier 1	0.380 (0.393)	0.561 (0.438)	-0.151 (0.634)	-0.283 (1.061)
Lag 2 Tier 1	-0.341 (0.230)	-0.222 (0.242)	-0.241 (0.778)	-0.084 (1.097)
Lag 3 Tier 1	-0.552** (0.232)	-0.474** (0.226)	-1.184 (0.829)	-1.350 (1.323)
<i>R</i> ²	0.744	0.759	0.462	0.452
<i>N</i>	770	700	770	700
<i>Counties</i>	70	70	70	70
<i>Controls</i>		Yes		Yes

Clustered standard errors in parentheses

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

Note: Table depicts estimates of equation 3. Standard errors are clustered by county. All columns include time and county fixed effects. Columns (2) and (4) include lags of the unemployment rate, income per capita, and population growth as additional controls. Columns (1) and (2) exclude from estimation of the treatment effects any counties designated as Tier 1 by an assignment rule besides the primary one based on distressed rank. Columns (3) and (4) label all Tier 1 counties as being treated and instrument for treatment with the distressed rank assignment rule.

randomization inference in a small neighborhood around the threshold.

We implement their approach for employment in table 7 and unemployment in table 8. 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 6. P-values are generated 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 in row one of table 7 implies that receiving the credits increases employment in a county by 2.5% after three years, similar to the parametric ITT results, though the estimate is not statistically significant. The estimates fall to 1.1% for a bandwidth of 10 ranks, and 0.5% 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 in row one of table 8 implies that receiving the credits decreases the unemployment rate in a county by 0.5 percentage points after three years, somewhat smaller in magnitude than the parametric estimates above. Expanding the window leads to more precisely estimated effects of -0.9 to -1.2 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.

It is interesting to contrast this with the diminution of the estimated effects on employment seen in table 7 at wider bandwidths. Expanding the window to include the most severely distressed counties biases these estimates downwards, indicating that the counterfactual trajectory for employment in these counties in the absence of the program was to fall further behind the control counties. The migration of unemployed workers out of these most severely distressed counties can reconcile these divergent bias patterns.

Table 7: Local estimates: Employment

Dependent Variable: Log Employment			
Window	1 Years	2 Years	3 Years
6 ranks	-0.006 [0.484]	0.005 [0.760]	0.025 [0.224]
N	61	55	49
Counties	25	23	21
10 ranks	-0.009 [0.209]	-0.003 [0.760]	0.011 [0.444]
N	119	108	97
Counties	29	29	28
20 ranks	-0.008 [0.182]	-0.008 [0.378]	0.005 [0.684]
N	200	181	161
Counties	44	44	43

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

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

Note: Table depicts the difference in mean outcomes for treated and control counties within a bandwidth of +/- 10 distress ranks around the policy threshold.

Table 8: Local estimates: Unemployment

Dependent Variable: Unemployment Rate			
Window	1 Years	2 Years	3 Years
6 ranks	-0.19 [0.50]	-0.43 [0.31]	-0.49 [0.33]
N	61	55	49
Counties	25	23	21
10 ranks	-0.18 [0.34]	-0.69** [0.02]	-0.89** [0.02]
N	119	108	97
Counties	29	29	28
20 ranks	-0.30* [0.06]	-0.78*** [0.00]	-1.20*** [0.00]
N	200	181	161
Counties	44	44	43

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

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

Note: Table depicts the difference in mean outcomes for treated and control counties within a bandwidth of +/- 10 distress ranks around the policy threshold.

5.4 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.³ Rows 1 and 2 of table 9 present RD estimates of equation 2 for aggregate employment within the targeted industries and non-targeted industries separately. Point estimates are imprecise, but show substantial increases in target industry employment as would be expected. Non-target industry employment levels are declining in treated counties relative to control counties, suggesting possible crowd-out effects.

Table 9: Regression discontinuity ITT estimates - Other outcomes

	1 yr later	2yrs later	3 yrs later
Log Employment Target Industries	0.038 (0.059)	0.066 (0.060)	0.064 (0.061)
Log Employment Non-Target Industries	-0.028 (0.021)	-0.033 (0.023)	-0.034 (0.025)
Log Hires Annual Total	-0.027 (0.033)	0.016 (0.041)	0.047 (0.058)
Log Separations Annual Total	-0.039 (0.028)	-0.034 (0.035)	0.015 (0.051)

Clustered standard errors in parentheses

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

Note: 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. N=2,670 for hires and separations. N=2,601 for industry employment due to censoring of industry level data in some small counties. 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.

As a robustness check, we re-estimate the RD specifications using Census QWI data on

³At 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).

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 3 and 4 in table 9 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, point estimates for separations are actually positive after three years, indicating that some churning of employment may have been induced by the program. If the program did increase the rate of worker turnover at eligible firms in order to generate more tax credits it would not alter the net employment impact of the program, though it would reduce the program's cost effectiveness.

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. We document the importance of accounting for unobservables and mean reversion which will bias difference in differences estimates. Our RD estimates for employment are noisy but generally show a boost from the program of around 2%. We find substantial impacts on unemployment, with treated counties, those whose firms were eligible for large hiring tax credits, experiencing about 1 percentage point lower unemployment rates than under a counterfactual program offering much smaller credits.

References

- Amior, Michael and Alan Manning (2015), "The Persistence of Local Joblessness." CEP Discussion Papers dp1357, Centre for Economic Performance, LSE.
- 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 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.

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.

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.

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.