

Do Spatially Targeted Redevelopment Incentives Work?

The Answer Depends on How You Ask the Question

Andrew Hanson and
Shawn M. Rohlin

MERCATUS WORKING PAPER

All studies in the Mercatus Working Paper series have followed a rigorous process of academic evaluation, including (except where otherwise noted) at least one double-blind peer review. Working Papers present an author's provisional findings, which, upon further consideration and revision, are likely to be republished in an academic journal. The opinions expressed in Mercatus Working Papers are the authors' and do not represent official positions of the Mercatus Center or George Mason University.



3434 Washington Blvd., 4th Floor, Arlington, Virginia 22201
www.mercatus.org

Andrew Hanson and Shawn M. Rohlin. "Do Spatially Targeted Redevelopment Incentives Work? The Answer Depends on How You Ask the Question." Mercatus Working Paper, Mercatus Center at George Mason University, Arlington, VA, 2017.

Abstract

We compare several common program evaluation techniques in evaluating the Empowerment Zone (EZ) program, a large US urban redevelopment program that consists primarily of tax credits and is run by the federal government. Studying the federal EZ program as a means to examine methodology is advantageous for several reasons. First, the federal program's application process generated a set of areas that were qualified and applied but that did not receive the program and thus created a comparison group that should not suffer from application bias. Second, the program had preapplication rules for which areas were considered, generating a rules-based group of comparison areas. Third, the program is uniform across areas, so program characteristics are not endogenous to local needs. Last, the geography of recipient boundaries is accounted for by census tract areas, as are comparison areas. We examine outcomes of the program under various models, including standard cross-section, difference-in-differences, triple difference, instrumental variables, and regression discontinuity. We construct comparison groups using several alternatives for each style of model, including trimming by propensity score. Our results generally show wide-ranging estimates of program effectiveness, with both positive and negative point estimates and a range of statistical significance. The most robust result suggests that EZs may have increased the number of firms in targeted areas in the short term, but the longer-term impact is less clear. We conclude that caution should be taken when interpreting the results of any one evaluation method as definitive, and we suggest that the effect of the EZ program on outcomes of interest is uncertain.

JEL codes: H25, H32, R51

Keywords: program evaluation, methodology, economic redevelopment, employment, firm location

Author Affiliation and Contact Information

Andrew Hanson
Marquette University
PO Box 1881
Milwaukee, WI 53201
andrew.r.hanson@marquette.edu

Shawn M. Rohlin
Kent State University
PO Box 5190
Kent, OH 44242
smrohlin@kent.edu

Copyright 2017 by Andrew Hanson, Shawn M. Rohlin, and the Mercatus Center at George Mason University

This paper can be accessed at <https://www.mercatus.org/publications/spatially-target-redevelopment-incentives>

Do Spatially Targeted Redevelopment Incentives Work?

The Answer Depends on How You Ask the Question

Andrew Hanson and Shawn M. Rohlin

I. Introduction

Spatially targeted, or place-based, incentive programs abound in the United States. Although the details of these programs can be as different as the areas they target, the common theme is that the programs confer benefits on the basis of geographic location within a homogeneous unit of geography. Often, program benefits come in the form of tax incentives (but also as grants and capital infusions and by other means), creating policy heterogeneity within an otherwise policy-homogeneous unit.

Place-based programs have been the subject of inquiry in a large and growing body of empirical studies by academics. Many of these studies focus on outcomes for residents or firms within the boundary created by the policy, and they examine wages, employment, poverty, property values, and business location. Neumark and Simpson (2015) provide an exhaustive review of the current state of the literature and point out that zone-based programs such as Enterprise Zones or Empowerment Zones receive most of the attention. Neumark and Simpson conclude that the evidence of the effectiveness of these programs is decidedly mixed. This view comes from a broad interpretation of the literature as a whole because individual studies rarely leave much room for ambiguity. Studies find either positive effects (in some cases quite large), as in Ham et al. (2011), Freedman (2013), and Busso, Gregory, and Kline (2013); or negative,

null, or diffuse effects as in Oakley and Tsao (2006), Elvery (2009), Hanson (2009), Neumark and Kolko (2010), and Reynolds and Rohlin (2015).¹

The goal of this paper is to explore the role that evaluation technique plays in contributing to the mixed findings in the literature on place-based policies. We compare several common program evaluation techniques, as outlined in Imbens and Wooldridge (2009), in evaluating the Empowerment Zone (EZ) program, a large US urban redevelopment program that consists primarily of tax credits. The federal EZ program is advantageous for our examination for several reasons: (a) it has a mostly uniform set of benefits that do not depend on local variation in characteristics, (b) the program generated a set of comparison areas through an application process and through rules-based criteria, and (c) the geography of targeted areas and application areas is well documented.

To understand how methodology might influence the mixed findings in the literature, we examine outcomes of the federal EZ program under cross-section, difference-in-differences, triple difference, instrumental variables, and regression discontinuity models. We construct various comparison groups using program rules, rejected applicants, qualified areas, and areas identified by propensity score estimation that are most similar to treated areas. We also examine several robustness checks within each method to assess the sensitivity of assumptions of particular methods. We apply all these evaluation methods to study the effect of the EZ program on the number of firms and the employment at firms within the EZ boundary, examining both a short-term effect and a longer-term effect of the program. The literature on the EZ program examines a range of different outcomes, including poverty and property values, but we focus on

¹ The hope is that by redeveloping specific parts of a city, state, or region, geographically targeted incentives will lead to positive spillover benefits in the larger area. Work by Hanson and Rohlin (2013) shows that, in fact, the spillover effect of the federal Empowerment Zone program was negative—targeted areas caused business relocation from surrounding and similar areas, with a net zero effect on the local economy.

firm-based effects in this paper. Examining firm-based outcomes matches with EZ program goals of providing economic opportunity to targeted areas, and it has the added advantage that the data are of higher frequency than other outcomes that use census data measured only once every 10 years.

We demonstrate that popular program evaluation methods, some of which are used in the existing literature, yield different results for the EZ program. Our analysis shows that difference-in-differences results generally produce positive outcomes, but these findings vary with the comparison group and usually are not precisely estimated. Triple difference estimation uniformly suggests that the program has a positive effect on employment and on the number of firms. This method is generally precisely estimated, whereas using instrumental variables within the differencing framework produces large, positive, but statistically imprecise, results. Finally, regression discontinuity designs, which are rarely used in the literature on the EZ program despite their recent prominence in program evaluation more generally, produce a wide range of findings that are sensitive to the use of the forcing variable, bandwidth, and control function. Our most robust finding across methodologies is that EZs increased the number of firms in designated areas in the short term, but even this finding varies across estimation methods.

Overall, these findings suggest that the choice of methodology is influential in determining the outcome of an evaluation of the federal EZ program. Some methods may be superior to others for evaluating a particular program, depending on the particular rules or environment surrounding the program, but a robust evaluation of spatially targeted policies should generally extend beyond one type of estimation strategy. We view the findings of this paper as a caution to program evaluators about using any single method or single evaluation as an input to benefit-cost analysis of spatially targeted incentive programs.

The remainder of the paper begins with a brief overview of place-based policy evaluations, with a focus on the primary concerns in estimation. The third section describes the details of the federal EZ program and describes why it is a good candidate to compare current program evaluation methodologies. The fourth section outlines each identification strategy separately and discusses the benefits and drawbacks of each. The fifth section discusses the data used in estimation. The sixth section summarizes the results across methodologies, and the final section offers concluding comments.

II. Overview of Place-Based Policy Evaluations

The goal of a policy evaluation is to determine what (if any) impact a policy has on a group of interest. The group of interest may not necessarily be the group targeted by the policy and, in fact, many evaluations examine spillover or unintended consequences of policies (Hanson and Rohlin 2013). Once the group of interest has been selected, the challenge is to determine what would have happened to that group in the absence of the policy. An ideal setting for the evaluator is for the program to be randomly assigned across a large number of similar individuals (or whatever the unit of analysis is) so that those not receiving the policy make a reasonable comparison group (in the absence of spillover effects between treated and comparison areas). In practice, random assignment is rare—policies are assigned on the basis of need, geographic location, individual circumstance, or other characteristics that may directly influence measuring outcomes of interest. This is especially true for place-based policies, where random assignment is nonexistent, treated areas are nearly always chosen because of characteristics, and spatial spillovers are likely, all of which make an unbiased evaluation challenging.

In the absence of random assignment, the choice for evaluators becomes how to construct a comparison group that will give the best representation of what would have happened to the

group of interest in the absence of the program. There is no general prescription for constructing a comparison group, but a few guiding principles give rise to standard methods:

- *Examine prepolicy characteristics and trends.* The observable characteristics of the group of interest should be similar to those of the comparison group before the policy takes effect. For place-based policies, this includes demographic and economic characteristics of residents as well as characteristics of the targeted area (for example, industry mix or regulatory climate). Whenever possible, one should also examine any trends in these characteristics leading up to the adoption of the policy to see if the policy is implemented during a cycle that may bias measuring the effect of the policy. Both prepolicy levels and trends should be as similar as possible between the group of interest and the control group to avoid biasing estimates of program outcomes. Although we control for both initial conditions and prepolicy trends in many of our research designs, note that it is impossible to control for unobservable differences that may be present between the group of interest and the comparison group. This issue is complicated by the fact that controlling for characteristics that are, in fact, outcomes from the program will bias program estimates on the outcome of interest (Angrist and Pischke 2009).
- *Minimize selection bias.* Programs may require that recipients apply to receive a policy. If this is the case, then evaluators need to consider whether the application process itself makes the group receiving a policy different from the comparison group. For example, if local areas apply for state-backed tax breaks, maybe only the well-managed local areas go through the application process. If this were true, then a measurement of the tax breaks would be biased because it would include both the tax breaks and the fact that these areas are better managed. When possible, a comparison group should be

chosen from the group of applicants, an option only when not all applicants receive the policy. Selection bias may also occur if program administrators choose applicants because they are either particularly needy or particularly likely to succeed or fail in the absence of the program.

- *Avoid policy spillovers.* Programs may have an effect outside the group of interest, even though that is not the intent of the policy. The key for evaluating the effect of the policy on the group of interest is to choose a comparison area that is not subject to spillovers from the policy. If a comparison group has a negative (positive) spillover from the policy, then the positive effect of the policy will be overestimated (underestimated) on the group of interest. It is not necessarily obvious what areas may be subject to spillovers from place-based policies, but certainly geographically close or adjacent areas would be suspect.² It could also be that economically similar areas are subject to spillovers from a policy, as Hanson and Rohlin (2013) find.
- *Consider external validity.* Nearly all program evaluation techniques offer an estimate of what is termed the *local average treatment effect* or the *treatment on treated effect* (Angrist and Pischke 2009). For example, the regression discontinuity design we implement in this paper uses only treated areas that are within a narrow range of unemployment rates, so the effects we estimate apply only to that group. Estimating a local average treatment effect means that the estimated policy effect may not be applicable outside of the sample where estimation occurs. The policy may have a much different effect on areas with much larger or smaller unemployment rates. It is also true

² Positive spillovers could result from agglomeration economies—external benefits to firms that are not in the targeted area that happen because of the increase in business activity in the local area. Negative spillovers could result if firms move from nontargeted areas into the targeted area, with no gains in productivity.

that only a small number of applicants received the EZ program, and we will estimate the effect on those areas, but the EZ may have a different effect on newer applicants or a different set of applicants.

The existing literature, as aptly summarized in Neumark and Simpson (2015), confronts these challenges in various ways and to various degrees; however, nowhere is there a rigorous cross-method evaluation of a place-based policy. We offer such an evaluation here.

III. The Federal Empowerment Zone Program

The Empowerment Zone program was passed into law as part of the Omnibus Budget Reconciliation Act of 1993 (OBRA 1993, P.L.103-66). This program enabled the federal government to offer tax incentives to employers located in parts of economically distressed areas. The US Department of Housing and Urban Development (HUD) designated three rural areas and parts of six cities as EZs. Those EZs were chosen from a group of applications made by state and local governments. Applications were considered for areas where at least 20 percent of the population lived in poverty and 6.3 percent were unemployed (US General Accounting Office 2004). From 78 nominees (Wallace 2004), the federal government awarded EZ status to parts of Atlanta, Baltimore, Chicago, Detroit, Philadelphia/Camden, and New York. Rural EZs, which we do not include in our analysis, were formed in the Kentucky Highlands, the Mississippi Delta, and the Rio Grande Valley in Texas. Zones were established as groups of census tracts.

Figure 1 (page 35) shows maps of both the New York and Chicago EZ areas, respectively. As shown in the figure, the EZs are relatively small portions of the cities by land area, and they generally overlap with what (at the time) were impoverished areas. For New York, the EZ covers much of Harlem and East Harlem and a portion of the South Bronx. In Chicago,

the EZ area covers the Douglas community on the city's south side and the west side area (the Garfield Park neighborhood). For the original urban EZs, \$100 million in the form of Social Services Block Grant funds accompanied the tax incentives. The largest component of the EZ program is the wage tax credit, which allows employers operating in the zone that hire residents of the zone to claim up to a \$3,000 tax credit per employee. Other tax incentives offered to firms operating in EZ-designated areas include (a) an increase in the amount of immediate expensing allowed, (b) postponement of capital gains reporting, (c) an increase in small business stock exclusion, and (d) temporary permission for state and local governments to operate outside the normal restriction on tax-exempt bonds offered on behalf of EZ businesses.

Many of the nominees that did not receive EZ status were awarded a "runner-up" award called Enterprise Communities (EC),³ a less generous overall package of assistance with a limited set of tax incentives. The biggest difference between EZs and ECs is that EC employers cannot claim the wage tax credit and EC zones were typically allowed only \$3 million in Social Services Block Grant funds.⁴

³ The runner-up group consists of parts of the following cities: Akron, Ohio; Albany, Georgia; Albany, New York; Albuquerque, New Mexico; Birmingham, Alabama; Boston, Massachusetts; Bridgeport, Connecticut; Buffalo, New York; Burlington, Vermont; Charleston, South Carolina; Charlotte, North Carolina; Cleveland, Ohio; Columbus, Ohio; Dallas, Texas; Denver, Colorado; Des Moines, Iowa; East St. Louis, Illinois; El Paso, Texas; Flint, Michigan; Harrisburg, Pennsylvania; Houston, Texas; Huntington, West Virginia; Indianapolis, Indiana; Ironton, Ohio; Jackson, Mississippi; Kansas City, Kansas; Kansas City, Missouri; Las Vegas, Nevada; Little Rock, Arkansas; Los Angeles, California; Louisville, Kentucky; Lowell, Massachusetts; Manchester, New Hampshire; Memphis, Tennessee; Miami, Florida; Milwaukee, Wisconsin; Minneapolis, Minnesota; Muskegon, Michigan; Nashville, Tennessee; New Haven, Connecticut; Newark, New Jersey; Newburgh, New York; Norfolk, Virginia; Oakland, California; Ogden, Utah; Oklahoma City, Oklahoma; Omaha, Nebraska; Phoenix, Arizona; Pittsburgh, Pennsylvania; Portland, Oregon; Providence, Rhode Island; Rochester, New York; San Antonio, Texas; San Diego, California; San Francisco, California; Seattle, Washington; Springfield, Illinois; Springfield, Massachusetts; St. Louis, Missouri; St. Paul, Minnesota; Tampa, Florida; Waco, Texas; Washington, DC; and Wilmington, Delaware.

⁴ The nominees in Boston; Oakland; Houston; Kansas City, Kansas; and Kansas City, Missouri, were designated as Enhanced Enterprise Communities (EEC). EEC status gave these communities a more generous allocation of grant funds than the standard Enterprise Communities received. Two nominees, Cleveland and Los Angeles, were awarded the status of Supplemental Empowerment Zone (SEZ) (US General Accounting Office 2004), which did not allow for all of the tax benefits of regular EZs but included more generous grants than regular EZs received (\$450 million for Los Angeles and \$177 million for Cleveland).

In many ways, EC areas form a natural comparison group for EZ areas. The EC areas were nominated to be EZs by local governments, went through the application process, and were deemed worthy of some form of assistance, so they may have some unobservable characteristics in common with EZ-designated areas—characteristics that are correlated with outcomes of interest that we would want to separate from program effects. Indeed, several evaluations of the EZ program use ECs of the set of all EZ applicants as a control group (Krupka and Noonan 2009; Hanson 2009; Busso, Gregory, and Kline 2013; Reynolds and Rohlin 2015). For comparison to EZ areas, figure 2 (page 36) shows maps for two EC areas, Los Angeles and Pittsburgh. The Los Angeles EC covers the south-central part of the city, including the Watts area and the Florence and Normandie intersection where rioting began in response to the 1992 Rodney King verdict. The Pittsburgh EC covers areas bordering the Allegheny, Ohio, and Monongahela rivers, including the South Side Flats area and the North Shore, and parts of the Strip District.

One of our methodologies follows the previous literature and uses EC areas as a way to construct a control group to create a counterfactual for what would have happened in EZ-designated areas if not for the EZ program. We also use the hard cutoff for poverty and unemployment limits in the program in a regression discontinuity design. Furthermore, as pointed out in Hanson (2009), EZs were designated as part of a contentious budget bill, suggesting that they may have been used as a political bargaining chip to gain a favorable vote. This possible motive is potentially advantageous from an identification standpoint because it means that at least some EZ areas may have been chosen not on the basis of a notion of future success or failure, but because of congressional representation. We use this theory to create an instrument for EZ designation: member representation (and number of terms) on the powerful US House Committee on Ways and Means.

IV. Methods of Identifying Program Effects

We focus on comparing methodology for identifying the effects of the federal EZ program on two outcomes: (a) the number of employees working at firms located in EZ areas and (b) the number of firms located in EZ areas.⁵ We use the number of employees and the number of firms as outcomes because they represent economic activity in the targeted area that we can measure with greater frequency than we can measure other outcomes, which are available only every 10 years through the census. Also, the outcomes we use represent the primary goal of the EZ program, to redevelop the local economy through employment and to spur economic activity within the targeted areas. We examine these outcomes in both the short (two-year) and long (six-year) time horizon, because the impact of any program may change over time as markets react and information about program benefits reaches more members of the targeted group.⁶

Cross-Section Regression Comparisons

Cross-section regression as an evaluation technique does not carry many advantages beyond simplicity, especially when data exist both before and after the program intervention and obvious control groups are available to make a difference-in-differences comparison. We present cross-section regression estimates as a technique mainly as a benchmark and to demonstrate how the choice of control group influences outcome measurement.

The basic cross-section regressions take the form

$$\ln(Y_i) = \alpha + \beta(D = 1 \text{ if } EZ)_i + \mathbf{X}'_i \boldsymbol{\delta} + \varepsilon. \quad (1)$$

⁵ Other research examines property values (Hanson 2009; Krupka and Noonan 2009; Busso, Gregory, and Kline 2013), rents (Busso, Gregory, and Kline 2013), neighborhood demographics (Krupka and Noonan 2009, Busso, Gregory, and Kline 2013), and wages (Busso, Gregory, and Kline 2013). Busso, Gregory, and Kline (2013) also estimate the impact of the federal EZ separately by place of residence and place of work.

⁶ One method that we do not use in our comparison that has become common in program evaluation is the synthetic control method. The synthetic control method is useful in cases in which there is only one treated unit, when a case-study approach would typically be the only form of evaluation. See Abadie, Diamond, and Hainmuller (2010) for a detailed explanation and application of the synthetic control method.

We estimate equation (1) one year after the program takes effect (1996) and five years after it takes effect (2000). Y_i is either the number of employees at firms or the number of firms in census tract i . We control for a series of characteristics in 1990 levels and the change between 1980 and 1990, denoted by X'_i . These control variables are (a) unemployment rate, (b) poverty rate, (c) percentage of nonwhite population, (d) percentage of female-headed families with children, (e) percentage of population age 25 or older with at least a college degree, (f) average age of housing stock (and this term squared), and (g) home ownership rate.

We restrict the comparison area for estimating (1) in several ways. First, we restrict the sample to include only EZ- or EC-designated census tracts. This restriction is our closest match to the primary specification used in Busso, Gregory, and Kline (2013), and it limits bias from unobservable factors that are correlated with an area that goes through the application process and that might confound program estimates but that leaves a small sample size. Second, we estimate (1) restricting the comparison group to only census tracts in metropolitan areas that meet the poverty and unemployment qualifications of being an EZ. This comparison group leaves open the possibility that bias with apply, but it limits confounding factors by using a sample of areas that were still qualified for the program, and it increases sample size. Third, we estimate (1) restricting the sample by a propensity scoring method, described in the next section of this paper.

In addition to the potentially viable comparison areas, we also estimate two versions of (1) that are intended to produce biased estimates. These include (a) estimating (1) without restriction on the data and using all census tracts identified as being in a metropolitan area in the United States as our control group, and (b) using only census tracts that border actual EZs as the control group. The all-tracts sample will be biased by both the application process and the fact

that EZ areas are generally more distressed than other census tracts. The border-tracts sample will be biased if the EZ program has an effect on other parts of the city that are close to EZ areas, or a spillover effect. In theory, a spillover effect might be negative if the EZ causes displacement from nearby areas, but it might be positive if the EZ results in strong local agglomeration economies. Examining the results of the biased estimates helps in understanding the direction of potential bias from other specifications and may help in understanding true program effects.

Propensity Score Trimmed Sample

We use propensity scoring to create an additional comparison group to investigate how this choice influences estimates of program effectiveness. The reasoning behind creating a comparison group this way is to find census tracts that are most similar to treated areas using several different observable characteristics. We do this by first estimating how several preprogram characteristics are correlated with program assignment, then by using the group of census tracts that is most similar to treated areas as a comparison group to measure outcomes. The hope is that a comparison group that is similar by observable characteristics will also share unobservable characteristics with EZ areas.

Construction of the propensity score comparison group first requires estimating the effect of observable characteristics on EZ adoption by the following equation:

$$EZ_i = \alpha + \mathbf{X}_i\boldsymbol{\delta} + \varepsilon. \quad (2)$$

The \mathbf{X} variables in this model are preprogram factors and political representation. The preprogram factors in equation (2) are identical to the control variables in equation (1). Political representation variables follow the observation in Wallace (2004) and Hanson (2009) that representatives on the Ways and Means Committee appeared to be successful in obtaining EZ status for parts of their districts. The political factors in equation (2) are a dummy variable for

area representation by a Ways and Means Committee member and a continuous variable measuring the time that the representative had served on the committee at the time of EZ designation.

The second step in constructing a comparison group using the propensity score method is to use the estimated δ coefficients and characteristics of census tracts pre-policy to create a likelihood (or propensity) that each tract in our sample was actually assigned an EZ. We then take the estimates of the likelihood that a census tract receives an EZ and trim the sample to include only census tracts in the top 1 percent of the propensity score distribution, creating a comparison group that is most similar to actual EZ areas by measurable factors.

Difference-in-Differences Comparisons

The federal EZ program provides several natural comparison areas for a standard difference-in-differences analysis. The idea behind a difference-in-differences comparison is to find areas that represent the trajectory of “what would have happened in EZ areas if not for the program.” Data across time and geographic areas are readily available, and the program makes clear designation of treatment areas and control groups.

Our basic difference-in-differences estimating equation is

$$\ln(Y_{i,t}) = \alpha + \beta_1(D = 1 \text{ if } EZ)_i + \beta_2(D = 1 \text{ if } After)_t + \beta_3(D = 1 \text{ if } EZ) \times (D = 1 \text{ if } After)_{i,t} + \mathbf{X}'_i \boldsymbol{\delta} + \varepsilon. \quad (3)$$

We estimate equation (3) for the same set of comparison areas that we outlined in the cross-section regression section. Three of these areas—the EC group, qualified areas, and the propensity score trimmed sample—have legitimate reasons for inclusion as a control group. We also estimate (3) using two samples that have obvious problems as a control sample: (a) all metropolitan area census tracts and (b) census tracts bordering EZ areas. This method is a

propensity score trimming method, as proposed in Crump et al. (2009), notably different from the common approach of matching treated and comparison units in some manner through a propensity score.

Triple Difference Comparisons

Standard difference-in-differences estimation suffers from bias if the cities that were designated EZs are on a different growth path than comparison cities. Note that this criticism is also valid if only the neighborhoods were on a different growth path. It seems plausible, especially given the small number of treated areas, that the group of EZ cities could, on average, have been subject to differential change in outcomes of interest even in the absence of the program. For example, city living becoming chic again in New York and Chicago, the Atlanta Olympics, and the ongoing Inner Harbor redevelopment in Baltimore all may have contributed to differential growth in those treated cities in the 1990s, even in the absence of EZ designation.

Triple difference estimation, in which program effects are a comparison of how EZ tracts fared relative to other tracts within their city and a comparison between EC tracts and other tracts in EC cities, eliminates general city-level improvement as a potential confounding factor. The triple difference specification is

$$\begin{aligned}
 \ln(Y_{i,t,c}) = & \alpha + \beta_1(D = 1 \text{ if } EZ \text{ or } EC)_i + \beta_2(D = 1 \text{ if } After)_t \\
 & + \beta_3(D = 1 \text{ if } EZ \text{ City})_c + \beta_4(D = 1 \text{ if } EZ \text{ or } EC) \times (D = 1 \text{ if } After)_{i,t} \\
 & + \beta_5(D = 1 \text{ if } After) \times (D = 1 \text{ if } EZ \text{ City})_{t,c} \\
 & + \beta_6(D = 1 \text{ if } EZ \text{ or } EC) \times (D = 1 \text{ if } EZ \text{ City})_{i,c} \\
 & + \beta_7(D = 1 \text{ if } EZ \text{ or } EC) \times (D = 1 \text{ if } After) \times (D = 1 \text{ if } EZ \text{ City})_{i,t,c} \\
 & + \mathbf{X}'_i \boldsymbol{\delta} + \varepsilon.
 \end{aligned} \tag{4}$$

The same set of comparison areas we used for the cross-section and difference-in-differences estimation cannot be used for the triple difference specification, because we now

must consider only areas within EZ and EC cities for differencing. We estimate equation (4) using three potentially legitimate control groups and one that should be subject to bias.

First, we estimate a standard triple difference between EZ areas and their own city with EC areas and their own city. Next, we reestimate the standard triple difference but exclude areas that border EZ areas because those excluded areas may be subject to spillovers. Third, we limit the sample to areas within EZ and EC cities that met program qualifications but were not part of the application. Finally, we limit the control group sample to only areas that border EZ and EC areas, in an attempt to show how this choice might produce biased estimates because of spillovers.

Instrumental Variables with Differencing

Triple difference estimates add a layer of sophistication to an EZ program evaluation that insulates program estimates from any confounding factors that differentially affected the group of EZ cities relative to comparison area cities. These estimates still leave open the possibility that EZ areas within cities were chosen for reasons that were directly related to future success or failure even in the absence of the program. In effect, this means that policy assignment may be endogenously determined by an attempt to either pick winners or pick losers. A potential way to provide an unbiased estimate in the face of policy endogeneity is to find an instrument for program assignment. An instrument requires an observable characteristic with strong correlation with EZ status but no direct correlation with outcomes of interest.

Wallace (2004) points out that the vote on EZ legislation was particularly contentious, a circumstance that may open up the possibility that political representation is a valid instrument for EZ designation. Wallace also points out that EZ designation was more likely for applicants that were at least partially covered by a representative on the Ways and Means Committee. We use area representation and the number of terms serving on the committee as a potential

instrument for EZ designation. This instrument is advantageous because these congressional districts only partially overlap with EZ areas, essentially grouping them into areas that were endogenously designated and designated because of the exogenous influence of a powerful member of Congress. Hanson (2009) applies this instrument in an evaluation of the EZ program on resident outcomes, and Hanson and Rohlin (2011) apply a similar method to evaluating the heterogeneous effect of EZs by firm age. The first-stage regression is the following:

$$EZ_i = \alpha + \beta_1(D = 1 \text{ if Ways and Means Rep})_i + \beta_2(\text{terms})_i + \mathbf{X}'_i \boldsymbol{\delta} + \varepsilon, \quad (5)$$

which leads to the second-stage estimation for the difference-in-differences specification:

$$\ln(Y_{i,t2} - Y_{i,t1}) = \alpha + \beta_1(\widehat{EZ})_{i,t} + \mathbf{X}'_i \boldsymbol{\delta} + \varepsilon, \quad (6)$$

and to the second-stage estimation for the triple difference specification:

$$\begin{aligned} \ln(Y_{i,t2} - Y_{i,t1}) = & \alpha + \beta_1(D = 1 \text{ if EZ or EC})_i + \beta_2(D = 1 \text{ if EZ City})_c \\ & + \beta_3(EZ \times \widehat{EZ} \text{ City})_{i,c} + \mathbf{X}'_i \boldsymbol{\delta} + \varepsilon. \end{aligned} \quad (7)$$

For ease of use, we take the difference across time in the dependent variable to estimate equations (6) and (7). The reliability of (6) and (7) as unbiased estimates of the EZ program hinges on the assumption that membership on the Ways and Means Committee in 1993 is not directly correlated with growth in employment or the number of firms in EZ areas between 1990 and 2000. The primary way in which Ways and Means Committee members can influence growth in their districts beyond the Empowerment Zone is through other forms of federal spending. Hanson (2009) explores the validity of the primary assumption in using this instrument, showing that while there is a large gap in the *level* of federal spending between Ways and Means member districts and other members' districts, there is little difference in the *growth* in federal spending in this period. It is certainly possible that using a different instrument would produce different estimates than those we find here. However, we do not believe that another

instrument would meet the exclusion restriction criteria as well as Ways and Means Committee membership.

Regression Discontinuity Design

The use of regression discontinuity estimation in program evaluation has become increasingly popular in recent years, especially since the elucidation of the method by Imbens and Lemieux (2008) and Lee and Lemieux (2010). The basic idea behind the method is to compare those who are treated by a policy with those who are not treated, on the basis of the use of discontinuous or “cutoff” points in program eligibility rules. The method is appealing if the associated discontinuities are created for reasons that are not related to the outcomes of interest, such as budgetary constraints or the use of round numbers. For the Empowerment Zone program, two cutoff points in eligibility were used—a minimum of 6.3 percent local unemployment and a minimum of 20 percent poverty rate. The regression discontinuity comparison chooses treated and control groups that have values very close to those cutoff points and examines how outcomes in those areas changed between groups that received an EZ and those that did not. The idea is that choosing control and treated groups that are close to the cutoff points minimizes differences in other characteristics between those groups.

We estimate the following regression discontinuity specification:

$$\ln(Y_{i,t2} - Y_{i,t1}) = \alpha + \beta_1(D = 1 \text{ if } EZ) + \sum_{j=1}^d [\varphi_{1,j}(v_i - cp_t)^j + \varphi_{2,j}EZ_i(v_i - cp_t)^j] + \varepsilon \text{ if } (cp_t - b) \leq v_i \leq (cp_t + b), \quad (8)$$

where the coefficient of interest is β_1 , and the terms inside the summation are the control function. We estimate equation (8) using both a linear control function ($d = 1$) and a quadratic control function ($d = 2$). The idea behind the control function is to hold constant the relationship between the outcome of interest and the forcing variable, so that the β_1 term

represents a discontinuity in this relationship. We also estimate (8) using both the national unemployment cutoff ($cp = 0.063$, and v representing tract-level unemployment) and the national poverty rate cutoff ($cp = 0.20$, and v representing tract-level poverty rate). We estimate each cutoff and control function across three different choices of bandwidth around the cutoff (measured by b). For the unemployment cutoff estimates, b takes the values 0.5, 1, and 2 percentage points, respectively. For the poverty cutoff estimates, b takes the values 2, 4, and 6 percentage points, respectively.

V. Data

Data on our outcomes of interest, employment at zone firms, and the number of zone firms come from the Dun and Bradstreet (D&B) Marketplace database. HUD used the D&B data for the official interim assessment of the EZ program (Hebert et al. 2001), although in HUD's assessment, the data are not applied to reliable program evaluation techniques. The data consist of the fourth-quarter survey from the years 1994, 1996, and 2000. The D&B data are aggregated at the zip code level. To map the zip code-level data on local establishments to census tracts, we use a correspondence to match the geography of the EZ- and EC-designated areas. The correspondence determines what percentage of each zip code lies in a given census tract and assigns that percentage of zip code employment or establishments to the census tract. Our list of EZ and EC census tracts was obtained through personal correspondence with HUD and is partially available through that department's website. Each EZ- or EC-designated area is made up of several census tracts. We treat census tracts as the unit of observation, but we cluster standard errors by county.⁷

⁷ Clustering standard errors is nearly equivalent to clustering by EZ/EC area. Two EZ areas were cross-county areas—the Philadelphia/Camden EZ and the New York City EZ. Although New York is all in one city boundary, each of the five boroughs represents a separate county.

We match the D&B data to data from both the 1990 and 1980 US censuses to create control variables for pretreatment levels and pretreatment trends in the previous decade. The 1990 census tracts match exactly with program boundaries, and a correspondence file is used to match 1980 boundaries. Table 1 (page 37) shows summary statistics for EZs and some comparison areas for both 1990 levels and 1980–1990 changes in control variables. As shown in the table, EZ areas had substantially higher poverty and unemployment rates than average census tracts in metropolitan areas in 1990. EZs were also places where more nonwhite, non-college-educated, female-headed households lived relative to the average census tract. EZ areas also had substantially older housing stock and drastically lower home ownership rates. Notably, although 1990 levels of all characteristics are quite different, the change in those characteristics between 1980 and 1990 was not as drastic for some characteristics. Column five displays a *p*-value from a difference in means test between EZ areas and all other metropolitan area census tracts for comparison.

Table 1 also shows that EZ areas were more similar to their comparison EC areas than they were to the group of all metropolitan area census tracts. Although EZ areas still have higher poverty and unemployment rates than EC areas, these differences are smaller and in some cases not statistically meaningful. Also notable is that although EZ areas have higher levels of poverty, the trend between 1980 and 1990 suggests that EC areas were actually outpacing EZ areas in increases in poverty. Column six shows a *p*-value for a difference in means test between EZ and the comparison EC areas for all the characteristics we measure. This column shows that although many of the differences between EC areas and EZ areas are smaller in magnitude than the difference that exists between EZ and all other tracts, the differences that exist are statistically different than zero.

VI. Results

Tables 2–7 report results of estimating the effect of the EZ program across the range of evaluation techniques we outlined earlier in this paper. The tables show results for both the number of employees in designated areas and the number of firms in those areas, and they offer both long-term (five-year) and short-term (one-year) estimates. We discuss the results here, organized by estimation technique.

Cross-Section Comparison Results

Table 2 (page 38) presents the cross-section estimation results. The table displays the β coefficient from equation (1) across the various outcome measures (rows) and comparison areas (columns). These results give the most positive estimate of how outcomes were different in EZ areas after the program had been in place for either one year (short-term) or five years (long-term). Starting from row one, column one (short-term effects on employment, using the EC comparison group), the results suggest that EZ areas had 25.7 percent *fewer* employees than comparison areas. The standard error on this estimate is large and includes zero, suggesting that although the point estimate is negative, the null hypothesis of zero program effect cannot be ruled out. For the three legitimate comparison groups (EC, qualified, and P-score), the estimates suggest a negative effect of the program on both jobs and the number of firms, except in one case. The point estimates suggest as much as a 37.6 percent loss in the number of firms when compared with EC areas. Standard errors are large relative to point estimates for the short-term employment outcome regardless of the choice of comparison group, but the estimates are statistically meaningful for the number of firms when comparing with EC areas and the P-score group.

These results highlight that in general, cross-section estimation suffers from a classic selection problem—targeted areas are likely targeted because they have overall lower levels of

economic activity. A simple cross-section comparison will not be able to account for initially different levels, but differencing methods do so by examining the change in activity from before the program to after it. Notably, the all-tracts and border-tracts comparisons generally show positive point estimates, suggesting that although EZ areas may be worse off than traditional comparison areas, they are generally better off than the average metropolitan area census tracts and the areas that are immediately adjacent to those census tracts after the program is in operation.

Difference-in-Differences Comparison Results

The standard difference-in-differences results, presented in table 3 (page 39), highlight the ambiguity in estimates of EZ success even for a popular and straightforward identification method. Comparing how EZ areas grew relative to various comparison groups mostly yields results that are not statistically different from zero, although the point estimates are large in some cases. Compared with EC-designated areas, EZs created 9.3 percent more employment in the short term, shrinking to 1.4 percent in the long term, with neither estimate statistically different from zero. Growth in the number of firms was positive relative to EC areas, at 6 percent in the short term, but again shrinking to near zero in the longer term and not statistically meaningful in either case.

Comparing EZ areas with other qualified areas does not produce consistent results, as point estimates in the short term are positive but insignificant (or only significant at the 10 percent level for firm growth), whereas some point estimates in the longer term are negative, but again insignificant. The propensity score comparison group produces the largest point estimates, showing growth in employment of 18.9 percent in the short term and 14.5 percent in the longer term, with growth in the number of firms of 15.4 percent in the short and 14.8 percent in the

longer term. Standard errors, however, are large on most of these estimates, and only the short-term firm growth results are statistically meaningful.

Comparing EZ areas with both all other census tracts and bordering census tracts shows an increase in the number of firms between 11 and 12 percent in the short term, and this result is statistically meaningful. These comparison areas are meant, however, to be unsuitable because they are either much different from EZ areas or subject to spillovers from the program. The fact that they produce positive estimates in some cases should be a warning that the choice of comparison group can create a false perception of program success.

Triple Difference Comparison Results

The triple difference comparison results portray a large, positive, and precisely estimated effect of the EZ program on both the number of employees in EZ areas and the number of firms in EZ areas in both the short term and longer term. These results, shown in table 4 (page 40), examine how EZ areas grew relative to their surrounding city and compare that with how comparison areas grew relative to their respective surrounding cities.

Comparing the relative growth between EC areas and their surrounding city, EZ areas increased employment by 27.2 percent in the short term and 30.7 percent in the long term. Both of these results are statistically meaningful at conventional levels and remain even after excluding border areas of the larger city (column two). This comparison group also shows the number of firms growing in EZ areas—by 14.4 percent in the short term and by 19.1 percent in the longer term. The positive, statistically meaningful result for firms is robust to excluding border areas as shown in column two.

Using only other qualified areas in the cities surrounding EZ and EC locations reduces the size of the estimates in columns one and two, and the estimates do not maintain statistical

precision at conventional levels. This suggests that although EZ areas did better than EC areas relative to their surrounding cities, the EZ areas did not always do better than other impoverished parts of their surrounding cities. Using only the border areas to difference with, we find that EZs were more successful, but these estimates should be biased if border areas are subject to spillovers, and they suggest that these spillovers may be negative.

Instrumental Variables Comparison Results

Table 5 (page 41) displays the instrumental variables regression results, where the EZ status of a census tract is treated as endogenous to future outcomes of employment and firm location and we use political representation to create exogenous variation in policy assignment. We display results where we instrument for EZ status in both the difference-in-differences and triple difference frameworks (where ECs are the comparison area in either case).

Both sets of results show large point estimates but also large standard errors. The size of the coefficients suggests large (up to 130 percent) increases in employment in the short term, and up to 90 percent increases in the number of firms in the long term. None of these results are statistically different from zero because the standard errors are large, leaving a high degree of ambiguity as to the true effect of the EZ program. We view the importance of these results as casting doubt on the standard triple difference results that show positive effects of the program. If the instrumental variables results are closer to the truth, our best guess is a very large effect from the program, but the results suggest that this estimate has a high degree of uncertainty.

Regression Discontinuity Results

Regression discontinuity results rely on two separate cutoff points: the unemployment and poverty thresholds, respectively. Table 6 (page 42) displays results using the unemployment threshold for the EZ program, and table 7 (page 43) shows results using the poverty threshold.

Although the method is identical, using the different program thresholds presents a vastly different set of results. Regression discontinuity has not previously been applied to evaluate the EZ program, but Freedman (2013) uses this method to evaluate a Texas state EZ-style program.

The unemployment threshold results compare places that had unemployment rates slightly lower than 6.3 percent in 1990 to places that had slightly higher rates and became EZ areas. The primary assumption in identifying the effects of the EZ program is that small differences in 1990 unemployment rates should not cause divergence in employment and firm location growth later in the decade (and that any growth differences are in fact the result of the EZ program). Table 6 shows that across nearly every specification, regardless of bandwidth choice or control function choice, the EZ had a large positive effect on both employment and the number of firms. The sizes of these estimates suggest that EZs were responsible for as much as a 240 percent increase in the number of employees in the short term and as much as 209 percent employment growth in the longer term. The employment results all have small standard errors and are statistically significant at the 1 percent level.

The results for the number of firms are nearly as strong. Point estimates suggest growth in the number of firms of up to 209 percent and at least 92 percent in the short term, statistically significant at conventional levels. The longer-term results for firm growth are less precisely estimated, but they still suggest that the EZ program caused a substantial increase in the number of firms in targeted areas.

Counter to these overwhelmingly positive results, the regression discontinuity results using the poverty cutoff produce several negative employment results, although the positive results for an increase in the number of firms in the short term remain. These results hinge on the assumption that areas with small initial differences in 1990 poverty rates will not have

differential growth in employment and firm growth in the 1990s (and that any differences are in fact caused by the EZ program).

Point estimates for the poverty cutoff regression discontinuity suggest highly uncertain effects of the EZ on employment in the short term—estimates range from an increase of 39.6 percent to a loss of 120 percent, although only one estimate is statistically meaningful at conventional levels. Point estimates for employment changes in the longer term are also highly variable, ranging from a loss of 22.9 percent to a gain of 52.8 percent, with most estimates achieving statistical significance at conventional levels. The poverty cutoff regression discontinuity estimates suggest positive effects on the number of firms in the short term, with effects ranging from 5 percent to as high as 99 percent increases; most of these are statistically significant at conventional levels. However, nearly all of the longer-term results for firms show a negative impact of the EZ program, with some estimates producing a statistically significant result.

Summarizing Program Results

To summarize the results in tables 2–7 in terms of the size of the impact they suggest and the certainty attached to the estimate (in the form of statistical significance), we present graphical plots in figures 3–6. Each figure plots the point estimate on the y -axis and a p -value for a test of statistical significance (where $p \leq 0.05$ is statistically significant as indicated by the dashed line). We separately present each outcome of interest (short-term employment, long-term employment, short-term number of firms, and long-term number of firms), excluding all cross-section estimates from table 2 and all unsuitable control groups, and displaying only results for the linear control function for regression discontinuity design estimates. One could view these figures as a summary of the results that one would potentially find across a range of studies on the EZ program that employ various identification methods.

Figure 3 (page 44) displays the summary of findings for employment effects of the EZ program in the short term. This meta-analysis shows that although all point estimates suggest a positive effect of the program, the size of the effect diverges greatly. Although some estimates produce statistically meaningful results, it seems most appropriate to describe the short-term effect of the EZ program on employment in targeted areas as uncertain. Examining across methods, regression discontinuity design results using the unemployment cutoff and most triple difference results show a statistically significant positive effect of the program, while the instrumental variables, regression discontinuity design using poverty cutoff, and difference-in-differences results have a range of point estimates but are never statistically significant.

Figure 4 (page 45) displays the summary findings for employment in the longer term. The longer-term employment effects are somewhat more promising in showing more estimates in the positive and statistically meaningful range, but these effects also show estimates in the negative and statistically meaningful range. Because nine estimates are in the negative or uncertain range and five are in the positive and significant range, we believe that an appropriate interpretation of the long-term effects of the EZ program on employment in EZ areas is, like the short-term effect, uncertain.

The most optimistic view on success of the EZ program is figure 5 (page 46), which shows the short-term effects on the number of firms operating in EZ areas. Figure 5 shows that a majority (nine) of our estimates suggest a positive and statistically meaningful effect of the EZ program on the number of firms operating in EZ areas. These effects are quite large, ranging from an increase of 20 percent to more than 200 percent, with all point estimates greater than zero. Some of the estimates suggest a higher degree of uncertainty, but even a few of the estimates below conventional statistical significance are statistically significant at the 10 percent level.

Unfortunately, the longer-term estimates of the EZ program on the number of firms are quite pessimistic. Figure 6 (page 47) shows that only 2 of the 14 estimates for the effect of the EZ program on the number of firms in the long term are positive and statistically significant at conventional levels. The longer-term estimates of the program include several negative estimates and many estimates with large standard errors, which suggests that any positive short-term effects of the program on the number of firms may not be long lasting.

Explaining Result Heterogeneity

The various methods we present in this paper produce a wide range of estimates for how the Empowerment Zone program affects firms and employment at firms located inside targeted zones. The variation in program estimates comes both across and within methods, but some methods offer more consistency than others do. All of the methods we demonstrate are common in the program evaluation literature, and with the exception of the cross-sectional analysis, all have a reasonable expectation of producing unbiased results of the EZ program. The question remains, “Why do these methods seemingly produce such a wide range of results?”

Part of the explanation could be that the EZ program itself produced a wide-ranging effect on local areas, or a heterogeneous treatment effect. Within the program, a wide range of cities—and neighborhoods within those cities—received benefits, and it is certainly possible that some of these areas flourished with the type of incentives offered under the EZ, whereas the incentives were not helpful in other areas. One could envision an area such as the New York EZ being ripe for economic redevelopment at the time, needing only a small push from the EZ incentives, while other zones such as the Detroit EZ would not succeed without a much larger investment.

If a heterogeneous treatment effect is present in the EZ program, this might help explain some of the variation across the regression discontinuity design estimates as the

bandwidth changes. In estimates that use only a narrow bandwidth, only a few treated areas (with the most favorable initial conditions) are considered treated; as the bandwidth increases, more treated areas are added that have less favorable initial conditions. In most of the regression discontinuity design estimates using the employment cutoff, we find larger point estimates using smaller bandwidth and smaller point estimates using larger bandwidth—the direction in which we would expect bias to work if more favorable initial conditions are correlated with a positive program effect.

Another explanation for heterogeneity across methods is that some methods are simply superior to others and that these estimates should be trusted, whereas other estimates are discounted. The triple difference methodology we employ produces the most consistent set of results (although not the largest in magnitude), and it has the advantage of using data from all EZ census tracts (compared with regression discontinuity design and instrumental variables estimates, which rely on subsets of tracts, or a local average treatment effect). A disadvantage of the triple difference estimates is that they would still be subject to bias if EZ locations were chosen for their promise of future success—a criterion that the application process evaluated.

Instrumental variables estimates attempt to eliminate the selection bias that may be present in triple difference comparisons, but those estimates rely on the assumption that the instrument (Ways and Means Committee representation) is not correlated with future economic success and had a strong correlation with receiving the program. Hanson (2009) demonstrates that the instrument is strong using conventional methods, but because it is uncorrelated with all other factors that drive future economic success, its accuracy is impossible to examine.⁸

Furthermore, the instrumental variables estimates are, like the regression discontinuity design

⁸ The standard first-stage F -statistic is 28.02, well over the acceptable threshold of 10; however, when standard errors are clustered at the city level, this first-stage F -statistic falls to 1.06.

estimates, local average treatment effects, and they rely on the subset of EZs that have overlap with Ways and Means member districts.

Overall, it seems appropriate to expect some variation across methodologies, and it is true that for some programs, some methods are objectively better than others. Although it is not possible to evaluate the source of all heterogeneity in our estimates of EZ program effects, we make a call to caution program evaluators to consider both within and across method variance in program estimates when making policy recommendations.

VII. Conclusions

We show that standard program evaluation methods produce a wide range of effects for the place-based federal Empowerment Zone program. Although there does seem to be a convergence across methods in supporting the conclusion that the EZ program caused a short-term increase in the number of firms in targeted areas, even this result shows a fair degree of variance in both size of effect and statistical significance. Even if the short-term effect of more firms is real, our results show that longer-term results for the number of firms in EZ areas are quite inconsistent; this finding suggests that any increase in the number of firms is short lived. Moreover, even if estimates showing positive benefits for the program are in fact correct, these benefits should be measured against negative spillover effects of the program found in Hanson and Rohlin (2013) and against program costs.

Our results show that for a program with uniform rules, in which several modern identification methods can be justified by reasonable assumptions, the methods do not converge on a result of program effectiveness. Although the methods presented here rely on different assumptions, none of the methods seem particularly arduous, so finding a high degree of variance across methods is somewhat surprising. Our findings suggest that program evaluators

should use a combination of techniques when determining program outcomes rather than relying on a single technique that may produce spurious results. Some methods may be more or less appropriate in different settings, and some of the methods we outline here may not be possible to use for all place-based policy evaluations. We suggest that any evaluation include a careful explanation of strengths and weaknesses of the chosen method and discussion of whether results are robust to alternative approaches.

Our findings do not allow us to suggest a method that is “best” for any one evaluation, but rather highlight that the choice of method can substantially influence results. We suggest that evaluators base the choice of method on the principle of finding the most realistic counterfactual for what happens to targeted areas—the best way of asking “What would have happened to the group of interest in the absence of the program?” We also suggest caution in using the findings of any single evaluation for policymaking, and we suggest that policymakers demand more rigor in the evaluation of programs before making decisions about them.

References

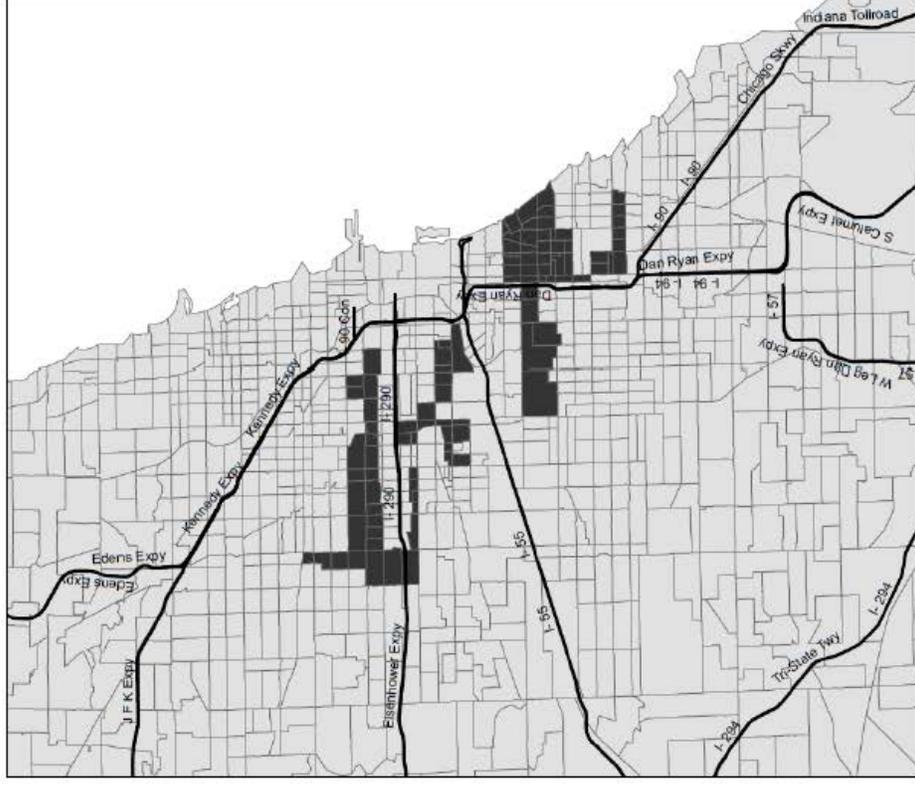
- Abadie, Alberto, Alexis Diamond, and Jens Hainmueller. 2010. "Synthetic Control Methods for Comparative Case Studies: Estimating the Effect of California's Tobacco Control Program." *Journal of the American Statistical Association* 105 (490): 493–505.
- Angrist, Joshua D., and Jörn-Steffen Pischke. 2009. *Mostly Harmless Econometrics: An Empiricist's Companion*. Princeton, NJ: Princeton University Press.
- Busso, Matias, Jesse Gregory, and Patrick M. Kline. 2013. "Assessing the Incidence and Efficiency of a Prominent Place Based Policy." *American Economic Review* 103 (2): 897–947.
- Crump, Richard K., V. Joseph Hotz, Guido W. Imbens, and Oscar A. Mitnik. 2009. "Dealing with Limited Overlap in the Estimation of Average Treatment Effects." *Biometrika* 96 (1): 187–99.
- Elvery, Joel. 2009. "The Impact of Enterprise Zones on Residential Employment: An Evaluation of the Enterprise Zone Programs of California and Florida." *Economic Development Quarterly* 23 (1): 44–59.
- Freedman, Matthew. 2013. "Targeted Business Incentives and Local Labor Markets." *Journal of Human Resources* 48 (2): 311–44.
- Ham, John C., Charles Swenson, Ayşe Imrohoroğlu, and Heonjae Song. 2011. "Government Programs Can Improve Local Labor Markets: Evidence from State Enterprise Zones, Federal Empowerment Zones and Federal Enterprise Communities." *Journal of Public Economics* 95 (7–8): 779–97.
- Hanson, Andrew. 2009. "Local Employment, Poverty, and Property Value Effects of Geographically-Targeted Tax Incentives: An Instrumental Variables Approach." *Regional Science and Urban Economics* 39 (6): 721–31.
- Hanson, Andrew, and Shawn Rohlin. 2011. "Do Location-Based Tax Incentives Attract New Business Establishments?" *Journal of Regional Science* 51 (3): 427–49.
- . 2013. "Do Spatially Targeted Redevelopment Programs Spillover?" *Regional Science and Urban Economics* 43 (1): 86–100.
- Hebert, Scott, Avid Vidal, Greg Mills, Franklin James, and Debbie Gruenstein. 2001. "Interim Assessment of the Empowerment Zones and Enterprise Communities Program: A Progress Report." Prepared for US Department of Housing and Urban Development, Washington, DC.
- Imbens, Guido, and Thomas Lemieux. 2008. "Regression Discontinuity Designs: A Guide to Practice." *Journal of Econometrics* 142 (2): 615–35.

- Imbens, Guido, and Jeffrey M. Wooldridge. 2009. "Recent Developments in the Econometrics of Program Evaluation." *Journal of Economic Literature* 47 (1): 5–86.
- Krupka, Douglas J., and Douglas S. Noonan. 2009. "Empowerment Zones, Neighborhood Change and Owner-Occupied Housing." *Regional Science and Urban Economics* 39 (4): 386–96.
- Lee, David S., and Thomas Lemieux. 2010. "Regression Discontinuity Designs in Economics." *Journal of Economic Literature* 48 (June 2010): 281–355.
- Neumark, David, and Jed Kolko. 2010. "Do Enterprise Zones Create Jobs? Evidence from California's Enterprise Zone Program." *Journal of Urban Economics* 68 (1): 1–19.
- Neumark, David, and Helen Simpson. 2015. "Place-Based Policies." In *Handbook of Regional and Urban Economics*, volume 5, edited by Gilles Duranton, J. Vernon Henderson, and William C. Strange, 1197–1287. Amsterdam: Elsevier.
- Oakley, Deirdre, and Hui-Shien Tsao. 2006. "A New Way of Revitalizing Distressed Urban Communities? Assessing the Impact of the Federal Empowerment Zone Program." *Journal of Urban Affairs* 28 (5): 443–71.
- Reynolds, C. Lockwood, and Shawn M. Rohlin. 2015. "The Effects of Location-Based Tax Policies on the Distribution of Household Income: Evidence from the Federal Empowerment Zone Program." *Journal of Urban Economics* 88 (July): 1–15.
- United States General Accounting Office (now the Government Accountability Office). 2004. *Federal Revitalization Programs Are Being Implemented, but Data on the Use of Tax Benefits Are Limited: Report to Congressional Committees*. GAO-04-306, Washington, DC.
- Wallace, Marc A. 2004. "Congressional Considerations and Urban Characteristics in the Selection of Empowerment Zones and Enterprise Communities." *Journal of Urban Affairs* 26 (5): 593–609.

Figure 1. Empowerment Zone Areas in New York and Chicago



New York



Chicago

Source: Author mapping of census tracts, provided by the Department of Housing and Urban Development, using ArcGIS software.

Table 1. Summary Statistics for Enterprise Zone and Comparison Areas

| | EZ (1) | All metropolitan (2) | EC (3) | Adjacent to EZ (4) | P-value (1)-(2) (5) | P-value (1)-(3) (6) | P-value (1)-(4) (7) |
|---|-----------|----------------------------|-----------|--------------------------|---------------------------|---------------------------|---------------------------|
| Poverty rate 1990 | 0.468 | 0.131 | 0.396 | 0.317 | 0.001 | 0.001 | 0.001 |
| Δ in poverty rate 1980–1990 | 0.057 | 0.041 | 0.067 | 0.028 | 0.064 | 0.203 | 0.004 |
| Unemployment rate 1990 | 0.229 | 0.067 | 0.164 | 0.160 | 0.001 | 0.001 | 0.001 |
| Δ in unemployment rate 1980–1990 | 0.049 | 0.014 | 0.040 | 0.020 | 0.001 | 0.186 | 0.001 |
| Percentage nonwhite 1990 | 0.856 | 0.199 | 0.671 | 0.678 | 0.001 | 0.001 | 0.001 |
| Δ in percentage nonwhite 1980–1990 | 0.041 | -0.1225 | 0.049 | 0.063 | 0.001 | 0.300 | 0.024 |
| Percentage with college degree 1990 | 0.070 | 0.215 | 0.089 | 0.138 | 0.001 | 0.001 | 0.001 |
| Δ in percentage with college degree 1980–1990 | 0.027 | 0.057 | 0.020 | 0.038 | 0.001 | 0.080 | 0.075 |
| Home ownership rate 1990 | 0.199 | 0.653 | 0.332 | 0.339 | 0.001 | 0.001 | 0.001 |
| Δ in home ownership rate 1980–1990 | 0.002 | -0.0178 | -0.01 | 0.009 | 0.001 | 0.058 | 0.410 |
| Percentage female household heads 1990 | 0.644 | 0.217 | 0.535 | 0.505 | 0.001 | 0.001 | 0.001 |
| Δ in percentage female household heads 1980–1990 | 0.081 | 0.067 | 0.066 | 0.071 | 0.117 | 0.212 | 0.479 |
| Average age of housing stock 1990 | 39.77 | 27.74 | 37.39 | 41.22 | 0.001 | 0.001 | 0.027 |
| Δ in average age of housing stock 1980–1990 | 3.47 | 5.07 | 4.72 | 4.82 | 0.001 | 0.001 | 0.003 |

Note: Geographically close areas are adjacent to EZ boundaries according to HUD definitions of these areas. *P*-values are from standard difference in means tests. Δ = change. EZ = Enterprise Zone.

Source: US Census Bureau, 1980, 1990.

Table 2. Cross-Section Identification Methods

| | vs. EC ^a | vs. Qualified | vs. P-score | vs. All tracts | vs. Border tracts |
|------------------------|---------------------|---------------|-------------|----------------|-------------------|
| Number of employees | | | | | |
| Short-term (1994–1996) | | | | | |
| | -0.257 | 0.036 | -0.179 | 0.308 | 0.038 |
| | (0.173) | (0.185) | (0.194) | (0.215) | (0.144) |
| N | 1,246 | 9,086 | 517 | 49,076 | 728 |
| Long-term (1994–2000) | | | | | |
| | -0.312 | -0.004 | -0.228 | 0.269 | 0.040 |
| | (0.199) | (0.205) | (0.214) | (0.237) | (0.165) |
| N | 1,304 | 9,436 | 536 | 51,702 | 728 |
| Number of firms | | | | | |
| Short-term (1994–1996) | | | | | |
| | -0.376** | -0.122 | -0.286** | 0.110 | 0.009 |
| | (0.156) | (0.145) | (0.116) | (0.169) | (0.072) |
| N | 1,246 | 9,086 | 517 | 49,076 | 728 |
| Long-term (1994–2000) | | | | | |
| | -0.376** | -0.136 | -0.313** | 0.088 | -0.003 |
| | (0.156) | (0.157) | (0.129) | (0.183) | (0.085) |
| N | 1,246 | 9,086 | 536 | 51,702 | 728 |

** = statistically significant at the 5% level.

Note: All estimates show beta coefficient from equation (1). “Number of employees” is the number of employees at firms operating within the census tract at firms in any industry. “Number of firms” is the total number of firms operating within the census tract. All regressions control for the following census tract characteristics in 1990 levels and in changes between 1980 and 1990: unemployment rate, poverty rate, percentage nonwhite population, percentage of female-headed families with children, percentage of population age 25 or older with at least a college degree, average age of housing stock (and this term squared), and the home ownership rate. Unit of observation is the census tract. All standard errors are clustered at the county level.

^a Enterprise Community (EC) comparison tracts include the Los Angeles and Cleveland Supplemental Empowerment Zones and the Washington, DC, Enterprise Community. The Los Angeles and Cleveland zones were awarded more generous allocations of block grants than the treated Empowerment Zones but were not allowed to claim the same tax advantages (notably, the wage tax credit). The Washington, DC, Enterprise Community was allowed to claim the wage tax credit beginning in August 1997. Results that exclude Los Angeles, Cleveland, and Washington, DC, show negative coefficients that are larger in magnitude and more precisely estimated than those presented here.

Source: All outcome data are from Dun and Bradstreet survey of firms zip code–level data aggregated to the census tract level using ArcGIS software. Census data from US Census Bureau, 1980, 1990.

Table 3. Difference-in-Differences Identification Methods

| | vs. EC ^a | vs. Qualified | vs. P-score | vs. All tracts | vs. Border tracts |
|------------------------|---------------------|---------------|-------------|----------------|-------------------|
| Number of employees | | | | | |
| Short-term (1994–1996) | | | | | |
| | 0.093 | 0.071 | 0.189* | 0.027 | 0.081* |
| | (0.107) | (0.097) | (0.103) | (0.094) | (0.039) |
| N | 2,550 | 18,059 | 1,040 | 98,596 | 1,456 |
| Long-term (1994–2000) | | | | | |
| | 0.014 | -0.040 | 0.145 | -0.185 | 0.076 |
| | (0.137) | (0.129) | (0.119) | (0.127) | (0.059) |
| N | 2,608 | 18,409 | 1,059 | 101,222 | 1,456 |
| Number of firms | | | | | |
| Short-term (1994–1996) | | | | | |
| | 0.060 | 0.093* | 0.154*** | 0.119** | 0.110*** |
| | (0.062) | (0.049) | (0.056) | (0.051) | (0.032) |
| N | 2,550 | 18,059 | 1,040 | 98,596 | 1,456 |
| Long-term (1994–2000) | | | | | |
| | 0.007 | -0.029 | 0.148 | -0.137 | 0.078* |
| | (0.010) | (0.092) | (0.107) | (0.093) | (0.036) |
| N | 2,608 | 18,409 | 1,059 | 101,222 | 1,456 |

* = statistically significant at the 10% level; ** = statistically significant at the 5% level.

Note: All estimates show beta₃ coefficient from equation (3). “Number of employees” is the number of employees at firms operating within the census tract at firms in any industry. “Number of firms” is the total number of firms operating within the census tract. All regressions control for the following census tract characteristics in 1990 levels and in changes between 1980 and 1990: unemployment rate, poverty rate, percentage nonwhite population, percentage of female-headed families with children, percentage of population age 25 or older with at least a college degree, average age of housing stock (and this term squared), and the home ownership rate. Unit of observation is the census tract. All standard errors are clustered at the county level.

^a Enterprise Community (EC) comparison tracts include the Los Angeles and Cleveland Supplemental Empowerment Zones and the Washington, DC, Enterprise Community. The Los Angeles and Cleveland zones were awarded more generous allocations of block grants than the treated Empowerment Zones but were not allowed to claim the same tax advantages (notably, the wage tax credit). The Washington, DC, Enterprise Community was allowed to claim the wage tax credit beginning in August 1997. Results that exclude Los Angeles, Cleveland, and Washington, DC, show coefficients that are nearly identical in magnitude not precisely estimated.

Source: All outcome data are from Dun and Bradstreet survey of firms zip code-level data aggregated to the census tract level using ArcGIS software. Census data from US Census Bureau, 1980, 1990.

Table 4. Triple Difference Identification Methods

| | vs. EC ^a (1) | vs. EC (exclude borders of both) (2) | vs. EC (both qualified within city) (3) | vs. EC (only borders of both) (4) |
|----------------------------|----------------------------|--|---|---|
| Number of employees | | | | |
| Short-term (1994–1996) | | | | |
| | 0.272** (0.107) | 0.310*** (0.119) | 0.197* (0.103) | 0.093* (0.053) |
| N | 23,584 | 19,775 | 8,561 | 7,005 |
| Long-term (1994–2000) | | | | |
| | 0.307** (0.119) | 0.351*** (0.132) | 0.180 (0.113) | 0.142*** (0.051) |
| N | 24,016 | 20,130 | 8,697 | 7,144 |
| Number of firms | | | | |
| Short-term (1994–1996) | | | | |
| | 0.144** (0.063) | 0.158** (0.067) | 0.101* (0.057) | 0.053 (0.043) |
| N | 23,584 | 19,775 | 8,561 | 7,005 |
| Long-term (1994–2000) | | | | |
| | 0.191** (0.082) | 0.217** (0.091) | 0.088 (0.077) | 0.101** (0.040) |
| N | 24,016 | 20,130 | 8,697 | 7,144 |

* = statistically significant at the 10% level; ** = statistically significant at the 5% level; *** = statistically significant at the 1% level.

Note: All estimates represent the beta₇ coefficient from equation (4). “Number of employees” is the number of employees at firms operating within the census tract at firms in any industry. “Number of firms” is the total number of firms operating within the census tract. All regressions control for the following census tract characteristics in 1990 levels and in changes between 1980 and 1990: unemployment rate, poverty rate, percentage nonwhite population, percentage of female-headed families with children, percentage of population age 25 or older with at least a college degree, average age of housing stock (and this term squared), and the home ownership rate. Unit of observation is the census tract. All standard errors are clustered at the county level.

^a Enterprise Community (EC) comparison tracts include the Los Angeles and Cleveland Supplemental Empowerment Zones and the Washington, DC, Enterprise Community. The Los Angeles and Cleveland zones were awarded more generous allocations of block grants than the treated Empowerment Zones but were not allowed to claim the same tax advantages (notably, the wage tax credit). The Washington, DC, Enterprise Community was allowed to claim the wage tax credit beginning in August 1997. Results that exclude Los Angeles, Cleveland, and Washington, DC, show coefficients that are slightly larger in magnitude precisely estimated at the one percent level in most cases.

Source: All outcome data are from Dun and Bradstreet survey of firms zip code–level data aggregated to the census tract level using ArcGIS software. Census data from US Census Bureau, 1980, 1990.

Table 5. Instrumental Variables Identification Methods

| | D-i-D with EC, instrumented | Triple Difference with EC, instrumented |
|------------------------|-----------------------------|---|
| Number of employees | | |
| Short-term (1994–1996) | | |
| | 1.300 | 1.297 |
| | (1.542) | (1.112) |
| N | 1,146 | 10,948 |
| Long-term (1994–2000) | | |
| | 0.849 | 1.067 |
| | (1.243) | (0.895) |
| N | 1,199 | 11,295 |
| Number of firms | | |
| Short-term (1994–1996) | | |
| | 0.107 | 0.420 |
| | (0.819) | (0.766) |
| N | 1,246 | 11,576 |
| Long-term (1994–2000) | | |
| | 0.802 | 0.911 |
| | (0.935) | (0.775) |
| N | 1,302 | 11,950 |

Note: D-i-D = difference-in-differences; DDD = triple difference. Estimates represent either the beta_1 coefficient in equation (6) in the case of D-i-D, or beta_3 coefficient in equation (7) for triple difference. Number of employees is the number of employees at firms operating within the census tract at firms in any industry. Number of firms is the total number of firms operating within the census tract. First stage of instrumental variables regression uses local representation on the Committee on Ways and Means and the number of terms the member had served at the time of EZ designation as instruments. All regressions control for the following census tract characteristics in 1990 levels and in changes between 1980 and 1990: unemployment rate, poverty rate, percentage nonwhite population, percentage of female-headed families with children, percentage of population age 25 or older with at least a college degree, average age of housing stock (and this term squared), and the home ownership rate. Unit of observation is the census tract. All standard errors are clustered at the county level.

^a Enterprise Community (EC) comparison tracts include the Los Angeles and Cleveland Supplemental Empowerment Zones and the Washington, DC, Enterprise Community. The Los Angeles and Cleveland zones were awarded more generous allocations of block grants than the treated Empowerment zones but were not allowed to claim the same tax advantages (notably, the wage tax credit). The Washington, DC, Enterprise Community was allowed to claim the wage tax credit beginning in August 1997. Results that exclude Los Angeles, Cleveland, and Washington, DC, show coefficients that are slightly larger in magnitude and not precisely estimated.

Source: All outcome data are from Dun and Bradstreet survey of firms zip code–level data aggregated to the census tract level using ArcGIS software. Census data from US Census Bureau, 1980, 1990.

Table 6. Unemployment Regression Discontinuity Identification Methods

| | Large bandwidth ($0.043 \leq b \leq 0.083$) | Medium bandwidth ($0.053 \leq b \leq 0.073$) | Small bandwidth ($0.058 \leq b \leq 0.068$) |
|------------------------|--|---|--|
| Number of employees | | | |
| Short-term (1994–1996) | | | |
| linear | 1.562*** (0.425) | 2.414*** (0.227) | 2.362*** (0.0308) |
| quadratic | 1.703*** (0.595) | 2.225*** (0.0318) | 2.371*** (0.0325) |
| N | 22,686 | 11,328 | 5,694 |
| Long-term (1994–2000) | | | |
| linear | 1.469*** (0.391) | 2.096*** (0.252) | 2.056*** (0.0317) |
| quadratic | 1.412** (0.560) | 1.901*** (0.0326) | 2.067*** (0.0324) |
| N | 24,022 | 11,998 | 6,004 |
| Number of firms | | | |
| Short-term (1994–1996) | | | |
| linear | 0.928** (0.368) | 1.048** (0.526) | 2.093*** (0.0296) |
| quadratic | 1.166** (0.516) | 2.199*** (0.227) | 2.098*** (0.0312) |
| N | 23,424 | 11,692 | 5,878 |
| Long-term (1994–2000) | | | |
| linear | 0.551* (0.318) | 0.703 (0.463) | 1.626*** (0.0310) |
| quadratic | 0.832* (0.466) | 1.671*** (0.343) | 1.632*** (0.0315) |
| N | 24,570 | 12,274 | 6,159 |

* = statistically significant at the 10% level; ** = statistically significant at the 5% level; *** = statistically significant at the 1% level.

Note: All estimates represent β_1 coefficient in equation (8) using the unemployment cutoff. “Number of employees” is the number of employees at firms operating within the census tract at firms in any industry. “Number of firms” is the total number of firms operating within the census tract. All regressions control for the following census tract characteristics in 1990 levels and in changes between 1980 and 1990: unemployment rate, poverty rate, percentage nonwhite population, percentage of female-headed families with children, percentage of population age 25 or older with at least a college degree, average age of housing stock (and this term squared), and the home ownership rate. Unit of observation is the census tract. All standard errors are clustered at the county level.

Source: All outcome data are from Dun and Bradstreet survey of firms zip code–level data aggregated to the census tract level using ArcGIS software. Census data from US Census Bureau, 1980, 1990.

Table 7. Poverty Regression Discontinuity Identification Methods

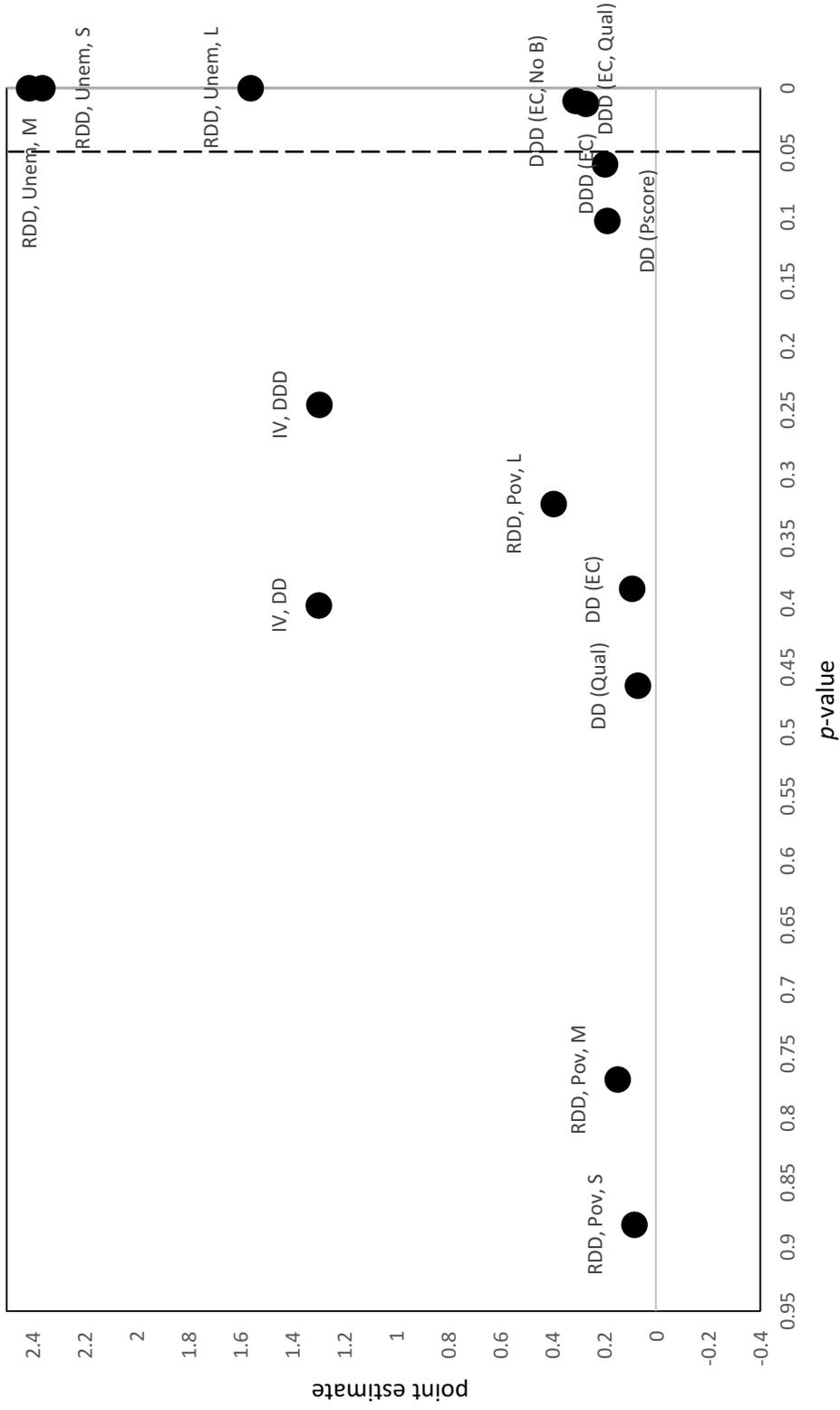
| | Large bandwidth ($0.14 \leq b \leq 0.26$) | Medium bandwidth ($0.16 \leq b \leq 0.24$) | Small bandwidth ($0.18 \leq b \leq 0.22$) |
|------------------------|--|---|--|
| Number of employees | | | |
| Short-term (1994–1996) | | | |
| linear | 0.396 (0.401) | 0.149 (0.510) | 0.084 (0.567) |
| quadratic | 0.598 (0.632) | -0.265 (0.577) | -1.213*** (0.046) |
| N | 12,018 | 7,764 | 3,788 |
| Long-term (1994–2000) | | | |
| linear | 0.0537 (0.524) | -0.229*** (0.033) | -0.214*** (0.035) |
| quadratic | 0.528** (0.249) | -0.190*** (0.036) | -0.210*** (0.039) |
| N | 12,763 | 8,258 | 3,996 |
| Number of firms | | | |
| Short-term (1994–1996) | | | |
| linear | 0.821*** (0.148) | 0.669*** (0.213) | 0.536** (0.212) |
| quadratic | 0.991*** (0.185) | 0.239** (0.117) | 0.054 (0.045) |
| N | 12,438 | 8,038 | 3,914 |
| Long-term (1994–2000) | | | |
| linear | -0.035 (0.169) | -0.213 (0.220) | -0.324 (0.224) |
| quadratic | 0.164 (0.197) | -0.596*** (0.149) | -0.835*** (0.044) |
| N | 13,087 | 8,452 | 4,085 |

* = statistically significant at the 10% level; ** = statistically significant at the 5% level; *** = statistically significant at the 1% level.

Note: All coefficients represent the beta_1 coefficient from equation (8) using the poverty cutoff. “Number of employees” is the number of employees at firms operating within the census tract at firms in any industry. “Number of firms” is the total number of firms operating within the census tract. All regressions control for the following census tract characteristics in 1990 levels and in changes between 1980 and 1990: unemployment rate, poverty rate, percentage nonwhite population, percentage of female-headed families with children, percentage of population age 25 or older with at least a college degree, average age of housing stock (and this term squared), and the home ownership rate. Unit of observation is the census tract. All standard errors are clustered at the county level.

Source: All outcome data are from Dun and Bradstreet survey of firms zip code–level data aggregated to the census tract level using ArcGIS software. Census data from US Census Bureau, 1980, 1990.

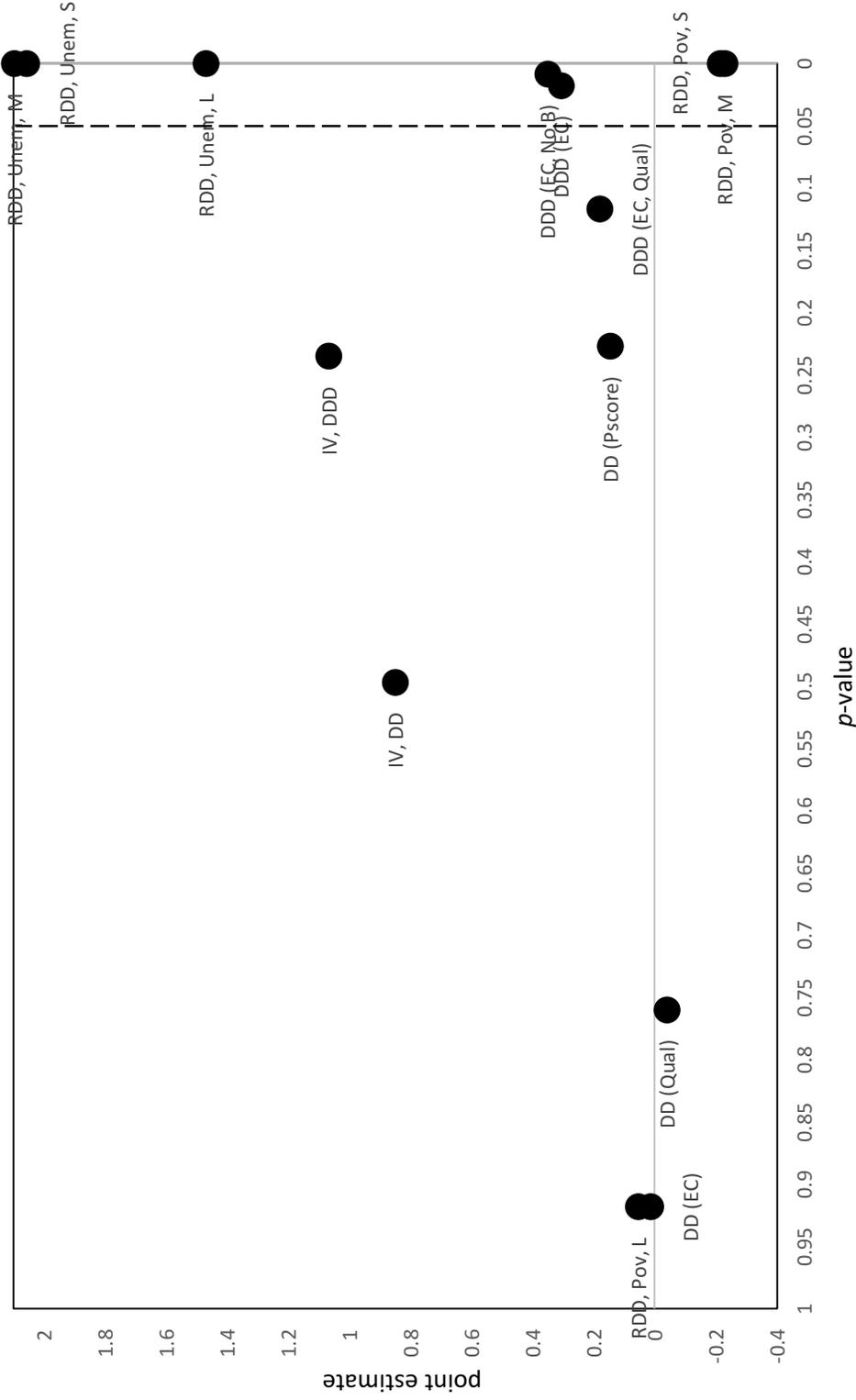
Figure 3. Summary of Results for Employment in Short Term



Note: EC = Enterprise Communities; DD = difference-in-differences; IV = instrumental variables; No B = that no border areas are part of the comparison group; Pov = RDD poverty threshold; Pscore = P-score-generated comparison group; Qual = qualified applicants as the comparison group; RDD = regression discontinuity design; S, M, and L = small, medium, and large bandwidths for RDD regressions; Unem = unemployment threshold.

Source: Authors' calculations.

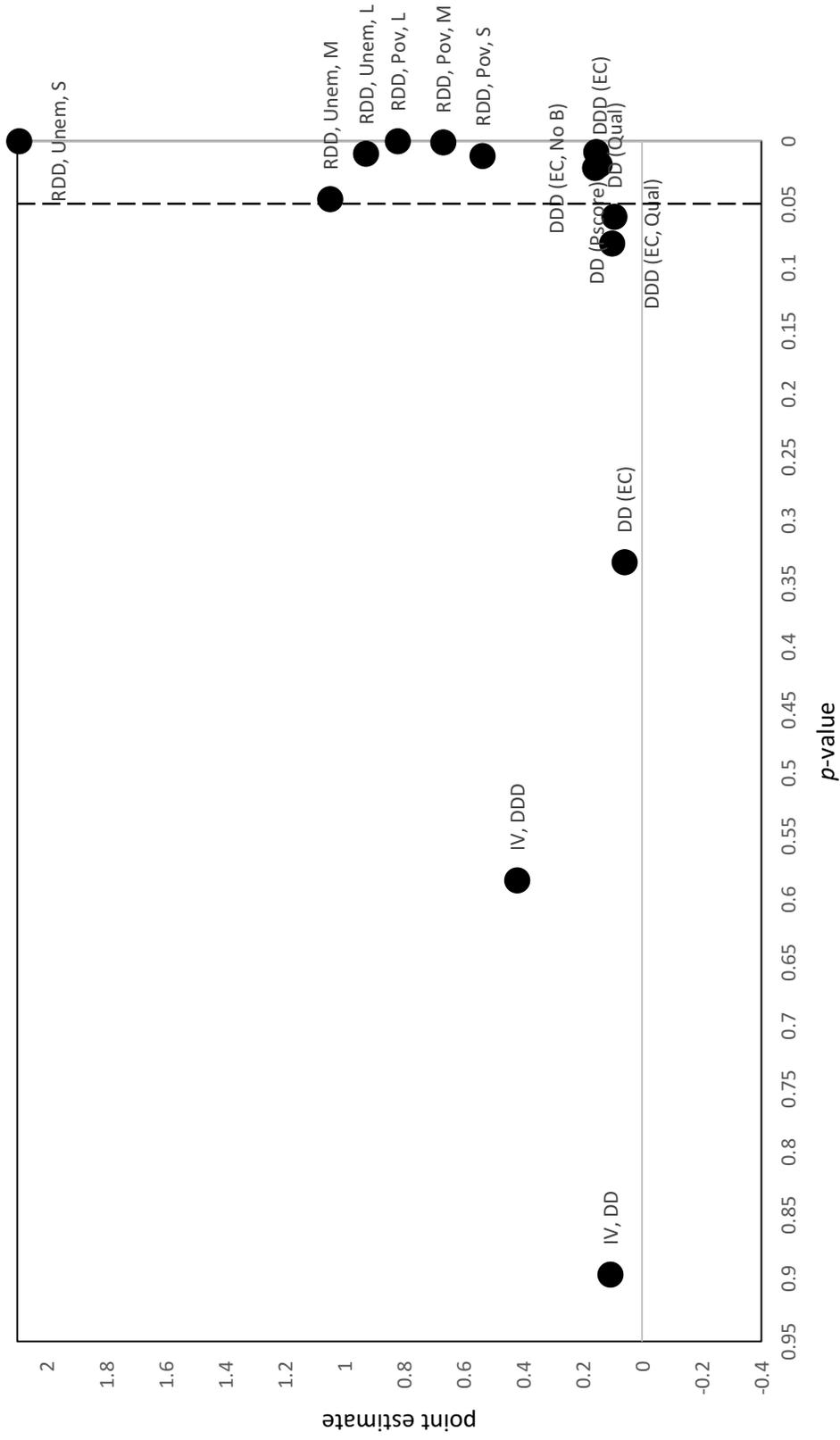
Figure 4. Summary of Results for Employment in Long Term



Note: EC = Enterprise Communities; DD = difference-in-differences; DDD = triple difference; No B = no border areas are part of the comparison group; Pov = RDD poverty threshold; Pscore = P-score-generated comparison group; Qual = qualified applicants as the comparison group; RDD = regression discontinuity design; S, M, and L = small, medium, and large bandwidths for RDD regressions; Unem = unemployment threshold.

Source: Authors' calculations.

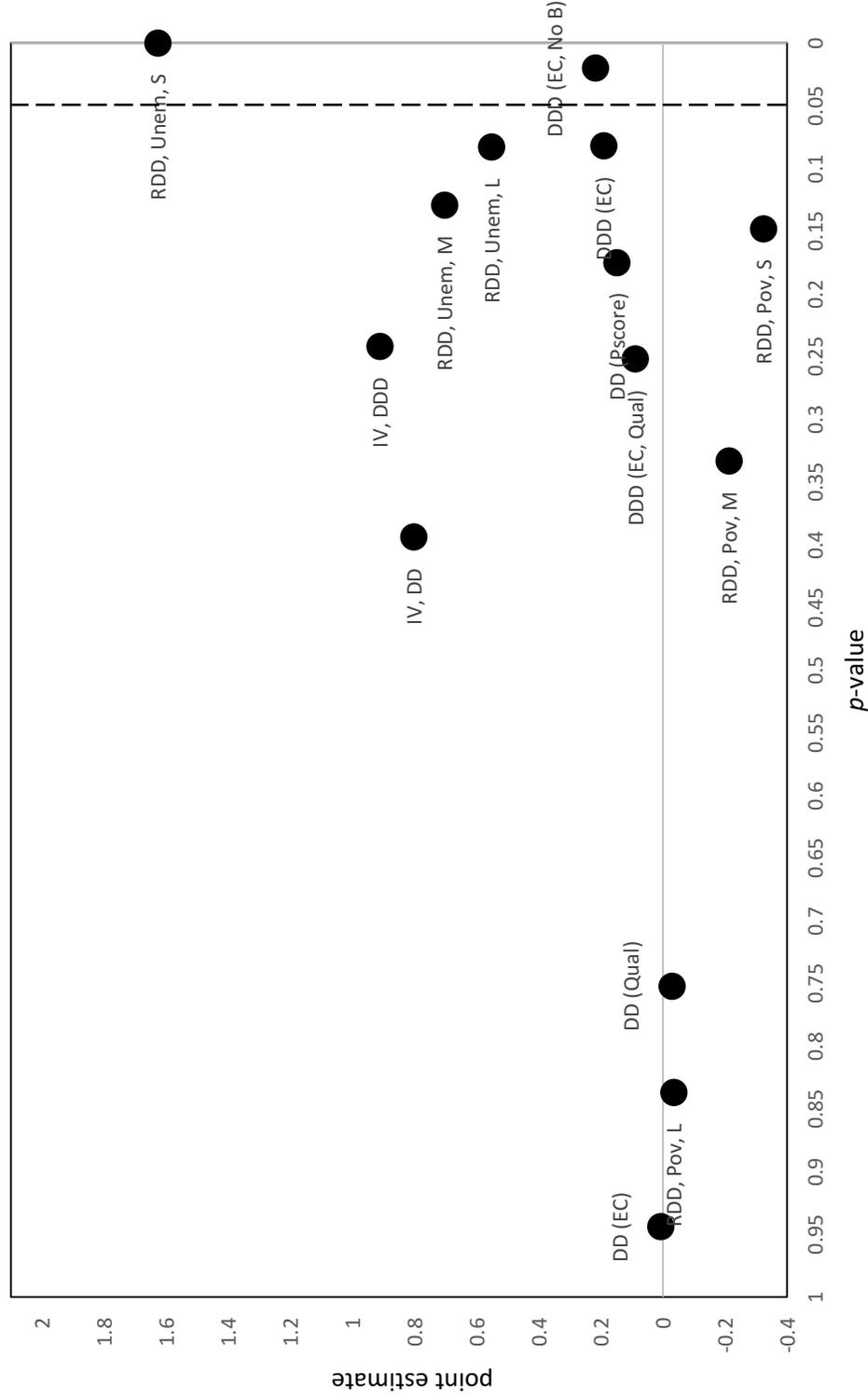
Figure 5. Summary of Results for Firms in Short Term



Note: EC = Enterprise Communities; DD = difference-in-differences; DDD = triple difference; IV = instrumental variables; No B = no border areas are part of the comparison group; Pov = RDD poverty threshold; Pscore = P-score-generated comparison group; Qual = qualified applicants as the comparison group; RDD = regression discontinuity design; S, M, and L = small, medium, and large bandwidths for RDD regressions; Unem = unemployment threshold.

Source: Authors' calculations.

Figure 6. Summary of Results for Firms in Long Term



Note: EC = Enterprise Communities; DD = difference-in-differences; DDD = triple difference; IV = instrumental variables; NoB = no border areas are part of the comparison group; Pov = RDD poverty threshold; Pscore = P-score-generated comparison group; Qual = qualified applicants as the comparison group; RDD = regression discontinuity design; S, M, and L = small, medium, and large bandwidths for RDD regressions; Unem = unemployment threshold.

Source: Authors' calculations.