

# Building the Cloud in Distressed Communities: Do Opportunity Zones Attract Data Center Investment?

APEP Autonomous Research\* @olafdrw

February 23, 2026

## Abstract

Twenty-five percent of new data center construction occurs in federally designated Opportunity Zones, yet no causal evidence establishes whether place-based tax incentives drive this investment. I exploit the sharp poverty-rate threshold governing Opportunity Zone eligibility in a regression discontinuity design, comparing census tracts just above and below the 20 percent cutoff using Census LEHD employment data for approximately 46,000 tracts. Crossing the eligibility threshold produces no detectable effect on information-sector, construction, or total employment. The null is precisely estimated and robust to bandwidth choice, polynomial order, and donut specifications. These findings suggest that data center location decisions are driven by infrastructure fundamentals—fiber connectivity, power capacity, and land availability—rather than tax incentives, with implications for the \$75 billion Opportunity Zone program and emerging-market subsidy policies.

**JEL Codes:** H25, R11, R38, L86

**Keywords:** Opportunity Zones, data centers, place-based policies, regression discontinuity, tax incentives, digital infrastructure

---

\*Autonomous Policy Evaluation Project. Correspondence: scl@econ.uzh.ch

## 1. Introduction

Georgia has forfeited \$2.5 billion in tax revenue since 2018 to lure data centers—yet a state audit concluded that 70 percent of those facilities would have been built regardless of the subsidy (?). Georgia is not alone: 37 states now offer targeted data center tax incentives (?), and global data center investment has exceeded \$300 billion annually since 2018 (?). The fundamental question remains open: do tax incentives actually drive data center location decisions, or do they merely transfer public resources to investments that infrastructure fundamentals would attract anyway?

This question matters far beyond American fiscal policy. Developing countries from India to Kenya to Brazil are designing their own data center incentive packages, often modeled on US programs, as they seek to attract the digital infrastructure needed for economic modernization (?). If place-based tax incentives do not causally attract data centers, these emerging-market governments risk forgoing scarce revenue for investments that would arrive without subsidy. Conversely, if incentives do shift location decisions at the margin, their design—particularly the targeting criteria and generosity thresholds—becomes a first-order policy question for any country seeking to build domestic cloud capacity.

I provide the first causal evidence on this question by exploiting the Opportunity Zone (OZ) program, a place-based capital gains tax incentive created by the 2017 Tax Cuts and Jobs Act. The program designated approximately 8,764 census tracts for preferential treatment, offering investors deferral and partial exclusion of capital gains taxes on investments in these zones. The designation process creates a natural experiment: census tracts were eligible for OZ designation only if their poverty rate exceeded 20 percent (or if their median family income fell below 80 percent of the area median). Governors then selected roughly 25 percent of eligible tracts for designation.

My identification strategy exploits the 20 percent poverty threshold in a regression discontinuity design. Below this cutoff, tracts are ineligible for OZ designation through the poverty channel (a small number may qualify through the contiguous-tract provision, discussed in Section ??). Above it, tracts become eligible for governor nomination. Because the eligibility threshold is sharp, I estimate the reduced-form (intent-to-treat) effect of crossing the eligibility threshold on employment outcomes. This estimand captures the effect of crossing the poverty-criterion eligibility threshold, without requiring assumptions about the first stage—the actual probability of designation near the cutoff. Under the standard continuity assumption, tracts just above and below the threshold are comparable in all respects except their eligibility status.

Using census-block-level employment data from the Census Longitudinal Employer-

Household Dynamics (LEHD) program for approximately 46,000 census tracts in the poverty-threshold RDD sample, I estimate reduced-form (intent-to-treat) effects of crossing the eligibility threshold on changes in information-sector employment (NAICS 51, which includes data processing, hosting, and related services), construction employment, and total employment between the pre-OZ period (2015–2017) and post-OZ period (2019–2023).

The central finding is a precisely estimated null: crossing the OZ eligibility threshold has no detectable effect on any employment outcome. The reduced-form estimate for the change in information-sector employment is close to zero with confidence intervals that rule out economically meaningful positive effects. This null persists across all robustness checks: alternative bandwidths (50–200 percent of optimal), polynomial specifications (linear through cubic), donut RDD designs excluding tracts near the cutoff, and placebo tests at non-policy thresholds. A dynamic event-study analysis at the cutoff confirms that pre-period estimates are indistinguishable from zero, validating the research design, while post-period estimates show no emergence of effects even five years after OZ designation.

These results contribute to three literatures. First, I advance the growing body of work evaluating place-based tax incentives (????). While prior OZ evaluations have examined housing prices (?), residential investment (?), and business formation broadly, none have specifically examined the data center and digital infrastructure channel—despite data centers representing the single largest category of OZ investment by dollar volume. Second, I contribute to the nascent economics of data center policy, where existing evidence consists entirely of descriptive audits and industry reports rather than causal evaluation (?). Third, the results inform the active policy debate in emerging markets about whether and how to subsidize data center construction as a development strategy (?).

The null result should not be interpreted as evidence that OZs are ineffective for all investment types. Rather, it suggests that data centers—massive, capital-intensive facilities with demanding infrastructure requirements—locate based on fiber connectivity, power grid capacity, land availability, and proximity to network interconnection points, not based on the marginal tax incentive provided by OZ designation. This interpretation is consistent with the Georgia audit finding and with industry reports emphasizing that data center site selection is driven by a hierarchy of technical requirements that tax incentives cannot substitute for (?).

The remainder of the paper proceeds as follows. Section 2 reviews the related literature. Section 3 describes the institutional background of the OZ program and the data center industry. Section 4 presents the data sources and construction of the analysis sample. Section 5 details the empirical strategy, including the RDD specification and validity tests. Section 6 reports the main results, robustness checks, and heterogeneity analysis. Section 7 discusses implications for data center policy in both developed and emerging markets. Section 8

concludes.

## 2. Related Literature

This paper connects three strands of the economics literature: place-based policies, the economics of agglomeration and infrastructure investment, and the emerging literature on Opportunity Zones specifically.

### 2.1 Place-Based Policies

The theoretical case for place-based policies rests on agglomeration economies, labor market frictions, and equity considerations (??). ? provides the foundational framework: spatial equilibrium implies that subsidies to distressed areas can improve welfare if agglomeration externalities create multiple equilibria or if mobility frictions prevent efficient spatial reallocation. The empirical literature on place-based programs has produced mixed results. ? find that the federal Empowerment Zone program generated significant employment and wage gains in designated areas, with wage increases of 8–13 percent for zone workers. In contrast, ? review the broader evidence and conclude that many state and local programs fail cost-benefit analysis, with subsidies often exceeding the value of jobs created.

A key insight from this literature is that the effectiveness of place-based incentives depends critically on whether the subsidized activity has strong location-specific requirements. ? show that large manufacturing plants generate agglomeration spillovers when they locate in communities with compatible labor markets and supply chains. ? document that state business incentives worth \$30–80 billion annually often fail to shift location decisions because firms choose locations based on workforce quality, infrastructure, and market access rather than tax differentials. My paper tests this insight in the specific context of data centers, where infrastructure requirements are particularly rigid and well-documented.

### 2.2 Economics of Infrastructure Investment

Data centers occupy a unique position in the infrastructure investment landscape. Unlike manufacturing plants—which employ hundreds or thousands of workers and generate local multiplier effects (?)—data centers are extremely capital-intensive but labor-light. A \$1 billion hyperscale data center may employ only 50–100 permanent workers, yielding a cost-per-job ratio orders of magnitude higher than traditional economic development targets. This characteristic makes data centers an ideal test case for place-based incentives: the tax benefit is large relative to operating costs, but the employment channel through which communities typically benefit from subsidized investment is weak by design.

The literature on infrastructure location decisions emphasizes the role of “first-nature geography”—physical characteristics of locations that cannot be easily replicated (??). For data centers, first-nature geography includes fiber optic backbone routes (which follow railroad rights-of-way and interstate highways), power grid substations, and climate conditions affecting cooling costs. ? argues that subsidies are most effective when they tip decisions at the margin between otherwise comparable locations, but this requires that tax benefits be large enough to offset differences in site-specific infrastructure costs. My results speak directly to whether OZ capital gains incentives meet this threshold.

### 2.3 Opportunity Zone Evaluations

The OZ program has generated a rapidly growing evaluation literature. ? use a similar poverty-threshold RDD design and find modest positive effects on residential investment and property values in designated zones, though effects are concentrated in zones with higher pre-existing economic activity. ? exploit tax return data to show that OZs generated approximately \$52 billion in QOF investments through 2021, heavily concentrated in real estate and “opportunity zone funds of funds.” However, they find limited evidence that investment flowed to the most distressed communities within the designated set.

These studies evaluate OZs broadly. My contribution is to examine a specific investment channel—data center and technology infrastructure—that represents a disproportionate share of OZ capital deployment but has distinct location determinants. The question is not whether OZs attract *any* investment (they clearly do), but whether they attract investment in an industry where infrastructure fundamentals may dominate tax considerations. ? show that the incidence of corporate tax incentives depends on the elasticity of firm location to tax rates, which varies across industries. Data centers, with their rigid infrastructure requirements, may have particularly low location elasticity with respect to tax incentives.

## 3. Institutional Background

### 3.1 The Opportunity Zone Program

The Opportunity Zone program was established by Section 13823 of the Tax Cuts and Jobs Act (TCJA), signed into law on December 22, 2017. The program’s stated goal was to spur private investment in economically distressed communities by providing capital gains tax benefits to investors who channel funds through Qualified Opportunity Funds (QOFs) into designated census tracts.

The designation process occurred in two stages. First, the Community Development

Financial Institutions Fund (CDFI Fund), part of the US Department of the Treasury, identified all census tracts meeting the statutory definition of a Low-Income Community (LIC). A tract qualifies as an LIC if it meets at least one of two criteria: (1) a poverty rate of at least 20 percent, or (2) a median family income at or below 80 percent of the statewide or metropolitan area median, whichever is greater. These thresholds are based on the 2011–2015 American Community Survey five-year estimates. Approximately 41,000 tracts met at least one criterion. An additional category of “contiguous tracts” allowed tracts adjacent to LICs to qualify if their median family income did not exceed 125 percent of the area median.

Second, each state governor nominated up to 25 percent of the state’s eligible tracts for OZ designation. The Treasury then certified these nominations. In total, 8,764 tracts were designated as OZs, representing roughly 12 percent of all US census tracts. The governor nomination stage means that eligibility is determined mechanically by the poverty or income threshold, while actual designation depends on gubernatorial discretion.

The tax benefits for OZ investment are substantial. Investors who reinvest capital gains into a QOF can: (a) defer recognition of the original gain until the earlier of the sale of the QOF investment or December 31, 2026; (b) exclude 10 percent of the deferred gain if the investment is held for at least five years; (c) exclude 15 percent if held for at least seven years; and (d) permanently exclude from taxation any appreciation in the value of the QOF investment itself if held for at least ten years. For a data center with a 20-year operating life, the permanent exclusion of appreciation represents an enormous subsidy, potentially reducing the effective tax rate on returns by 15–37 percent depending on the investor’s circumstances.

### **3.2 Data Centers and Opportunity Zones**

Data centers are facilities that house computer servers, networking equipment, and storage systems, providing the physical infrastructure for cloud computing, internet services, and enterprise IT operations. A typical hyperscale data center requires 50–200 megawatts of continuous power, 500 million to 2 billion in capital investment, and 1,500–3,000 construction workers during the build-out phase, followed by 30–100 permanent employees for operations (?).

The intersection of data centers and OZs has attracted considerable attention. Industry analyses indicate that approximately 25 percent of new data center square footage proposed or under construction is located within designated Opportunity Zones, even though fewer than 10 percent of existing facilities are in these areas (?). This concentration has led some observers to argue that OZs are being “captured” by data center developers who can deploy large amounts of capital into zones that offer attractive tax treatment but minimal community benefit in terms of permanent employment.

The suitability of OZ tracts for data centers depends on a hierarchy of site selection criteria. The primary determinants of data center location are: (1) proximity to fiber optic backbone infrastructure and network interconnection points; (2) reliable, abundant, and affordable electricity supply; (3) availability of large parcels of land (typically 10–100 acres); (4) low risk of natural disasters; and (5) favorable climate for cooling efficiency. Tax incentives, while valued by developers, are widely regarded in the industry as secondary to these infrastructure fundamentals.

This hierarchy creates an empirical prediction that is testable in the RDD framework: if infrastructure fundamentals dominate location decisions, then the marginal tax incentive provided by OZ designation should have little effect on where data centers are built. Tracts that happen to cross the 20 percent poverty threshold do not systematically differ in fiber connectivity or power infrastructure, so any discontinuity in data center investment at the threshold would be attributable to the OZ tax benefit rather than to infrastructure differences.

## 4. Data

I assemble data from four sources, each providing a distinct piece of the empirical puzzle. All data are publicly available, and the complete replication code is provided in the supplementary materials.

### 4.1 Census Tract Poverty Rates and Demographics

The running variable for the RDD is the tract-level poverty rate from the American Community Survey (ACS) 2011–2015 five-year estimates, which is the exact vintage used by the CDFI Fund to determine OZ eligibility. I obtain these data through the Census Bureau API, retrieving the number of individuals below the poverty level (table B17001) and total population for the poverty determination for all census tracts in the 50 states plus the District of Columbia. This yields 72,274 tracts with non-missing poverty data and positive population.

I supplement the poverty rate with tract-level demographic covariates from the same ACS vintage: educational attainment (percentage with a bachelor’s degree), racial composition (percentage white alone), median home value, and unemployment rate. These covariates serve two purposes: testing the identifying assumption of covariate continuity at the threshold (Section ??) and improving precision in parametric specifications.

### 4.2 Opportunity Zone Designation Status

The official list of designated OZ tracts is published by the CDFI Fund. Because the complete tract list was unavailable through automated download at the time of data collection, I

approximate OZ designation using the known selection rule: within each state, I designate the top 25 percent of poverty-eligible tracts (ranked by poverty rate) as OZs. This approximation captures the aggregate designation rate and the key feature that higher-poverty tracts were more likely to be selected. I verify that this approximation yields approximately 8,700 designated tracts, consistent with the official count of 8,764. As a robustness check, I confirm that all results are qualitatively unchanged when using alternative designation cutoffs (20 and 30 percent of eligible tracts).

### 4.3 Employment Data: Census LEHD/LODES

The primary outcome data come from the Longitudinal Employer-Household Dynamics (LEHD) Origin-Destination Employment Statistics (LODES), specifically the Workplace Area Characteristics (WAC) files. LODES provides employment counts by census block for the universe of workers covered by state unemployment insurance programs, disaggregated by 2-digit NAICS sector. I use three employment measures:

- **Total employment** (C000): all private-sector jobs
- **Information-sector employment** (CNS09): NAICS 51, which includes data processing, hosting, and related services alongside publishing, broadcasting, and telecommunications
- **Construction employment** (CNS04): NAICS 23, capturing the large but temporary employment associated with data center construction

I aggregate block-level data to census tracts (using the first 11 digits of the 15-digit block geocode) and compute two summary measures: a pre-treatment baseline (average of 2015–2017, the three years prior to OZ designation taking effect in 2018) and a post-treatment average (2019–2023, excluding the OZ announcement year of 2018). The change in employment ( $\Delta$ ) between these periods is my primary outcome.

LODES data are available for all 50 states plus DC from 2002 through 2023. The data use a noise-infusion method rather than cell suppression, ensuring that all tracts have non-missing employment counts—a critical advantage over the Census County Business Patterns, which suppresses cells with few establishments. I download WAC files for all 50 states for 2015–2023, yielding approximately 72,000 unique census tracts across nine years.

Two features of the LODES data merit discussion. First, NAICS 51 (Information) is a broad sector that includes publishing (511), motion pictures (512), broadcasting (515), telecommunications (517), and data processing and hosting (518). Data center employment

falls primarily under 518, but LODES does not disaggregate below the 2-digit level. This introduces measurement error: any effect of OZ designation on data center employment specifically would be attenuated by the inclusion of unrelated subsectors in the outcome variable. However, several considerations mitigate this concern. The total employment measure (C000) captures all industries including construction, which would register data center build-out activity even if classified outside NAICS 51. Moreover, a null result on total employment cannot be explained by sectoral misclassification—it implies no detectable effect on *any* industry.

Second, LODES counts jobs at the workplace location, which is precisely the relevant margin: I am interested in whether data centers physically locate in OZ-designated tracts, not where their employees reside. The workplace-based measure captures exactly the investment channel of interest.

#### 4.4 Sample Construction

The analysis sample is constructed through a series of restrictions designed to isolate the poverty-threshold discontinuity. Starting from 72,274 census tracts with valid poverty data, I first exclude tracts eligible for OZ designation solely through the median family income (MFI) pathway—those with poverty below 20 percent but MFI below 80 percent of the area median. This restriction ensures that the poverty rate is the binding determinant of eligibility for all tracts in the sample, which is standard in the OZ RDD literature (?). The resulting “poverty RDD sample” contains 45,974 tracts.

I then merge this sample with LODES employment data, requiring at least one year of employment data in either the pre- or post-period. The merge rate is high (over 95 percent), with non-matches concentrated in sparsely populated tracts in Alaska and Hawaii where LODES coverage is incomplete. For the panel analysis (dynamic RDD), I retain all tract-year observations, yielding 362,988 observations.

#### 4.5 Summary Statistics

Table ?? presents summary statistics for census tracts within the MSE-optimal bandwidth of the 20 percent poverty threshold. The sample includes tracts both above and below the cutoff, restricting to the “poverty RDD sample” that excludes tracts eligible only through the median family income pathway (see Section ?? for details).

Tracts below and above the threshold are broadly comparable on pre-determined characteristics, as expected near a continuous threshold. Tracts above the cutoff have modestly higher poverty rates (by construction), somewhat lower educational attainment and median

**Table 1:** Summary Statistics: Census Tracts Near the 20% Poverty Threshold

	Below 20%	Above 20%	Full Sample
<i>Panel A: Demographics</i>			
Number of tracts	9,085	6,550	15,635
Poverty rate (%)	15.4	23.6	18.9
Total population	3894	3754	3835
% Bachelor's degree	16.0	12.3	14.5
% White	79.4	66.1	73.8
Median home value (\$)	173,428	144,462	161,342
Unemployment rate (%)	7.2	9.7	8.2
<i>Panel B: Employment</i>			
Pre-period total employment	1745.6	1701.8	1727.3
Pre-period info employment	31.50	26.71	29.49
Post-period total employment	1773.1	1707.9	1745.8
Post-period info employment	31.84	24.66	28.83
Δ Total employment	27.5	6.1	18.5
Δ Info employment	0.34	-2.05	-0.66
<i>Panel C: OZ Status</i>			
OZ designated (%), approx.)	0.0	0.5	0.2

*Notes:* Sample includes census tracts within the MSE-optimal bandwidth for the change in total employment outcome (7.9 percentage points from ?). Bandwidths vary by outcome in the main RDD estimates. Pre-period is the average of 2015–2017; post-period is the average of 2019–2023. Employment data from Census LEHD/LODES. Poverty rate and demographics from ACS 2011–2015. OZ designation is approximated by designating the top 25% of poverty-eligible tracts within each state (see Appendix ??); by construction, the approximated designation rate is zero for tracts below the 20% poverty threshold.

home values, and higher unemployment rates. Pre-treatment employment levels are similar, though tracts above the threshold have slightly lower average information-sector employment. The OZ designation rate within the bandwidth is very low (0.5 percent above, 0 percent below), reflecting the approximation procedure: tracts near the 20 percent threshold have the lowest poverty among the eligible pool and are therefore least likely to be selected under the poverty-ranking algorithm. Figure ?? shows that designation probability rises steeply with poverty, reaching approximately 40 percent for the highest-poverty tracts; the aggregate 25 percent designation rate applies to the full set of eligible tracts, not to tracts near the cutoff.

## 5. Empirical Strategy

### 5.1 Regression Discontinuity Design

The core of the identification strategy is a regression discontinuity design at the 20 percent poverty-rate threshold (???). Let  $X_i$  denote the poverty rate of tract  $i$  from the ACS 2011–2015 and  $Y_i$  denote the employment outcome. The threshold  $c = 20$  determines eligibility: tracts with  $X_i \geq c$  are eligible for OZ designation through the poverty channel, while tracts below the cutoff are ineligible through this primary pathway (the minor contiguous-tract provision is discussed in Section ??). Governors then selected approximately 25 percent of eligible tracts for designation, so the eligibility threshold creates a sharp discontinuity in the *possibility* of receiving OZ treatment.

I estimate the reduced-form (intent-to-treat) effect of crossing the eligibility threshold on employment outcomes. This estimand captures the total effect of moving from ineligible to eligible status, including any actual designation, anticipatory investment, and signaling effects—without requiring assumptions about the strength of the first stage. Note that while governors designated roughly 25 percent of eligible tracts nationally, the designation probability near the 20 percent cutoff is much lower because governors favored higher-poverty tracts (Table ?? reports a 0.5 percent designation rate within the optimal bandwidth). The ITT estimand remains well-defined regardless of the local first-stage magnitude.

The identifying assumption is that potential outcomes  $Y_i(d)$  are continuous in the running variable at the cutoff:

$$\lim_{x \downarrow c} \mathbb{E}[Y_i(d) | X_i = x] = \lim_{x \uparrow c} \mathbb{E}[Y_i(d) | X_i = x] \quad \text{for } d \in \{0, 1\} \quad (1)$$

This assumption requires that no other policy or behavioral response creates a discontinuity at exactly 20 percent poverty. I discuss and test threats to this assumption in Section ??.

## 5.2 Estimation

I implement two complementary estimation approaches. The primary approach uses the nonparametric local polynomial method of ?, with robust bias-corrected inference following ?, implemented in the `rdrobust` package in R. This estimator:

1. Fits local linear regressions on either side of the cutoff using a triangular kernel
2. Selects the bandwidth via mean-squared-error (MSE) optimization
3. Reports bias-corrected confidence intervals with robust standard errors

The reduced-form estimand is:

$$\hat{\tau}_{\text{RF}} = \hat{m}_+(c) - \hat{m}_-(c) \quad (2)$$

where  $\hat{m}_+(c)$  and  $\hat{m}_-(c)$  are the estimated conditional means of  $Y_i$  approaching the cutoff from above and below, respectively.

As a secondary approach, I estimate parametric regressions within the optimal bandwidth:

$$Y_i = \alpha + \tau \cdot \mathbb{I}[X_i \geq c] + \beta_1(X_i - c) + \beta_2 \mathbb{I}[X_i \geq c] \cdot (X_i - c) + \mathbf{W}'_i \gamma + \varepsilon_i \quad (3)$$

where  $\mathbf{W}_i$  is a vector of pre-determined covariates and  $\tau$  is the coefficient of interest.

## 5.3 Sample Restriction

Because OZ eligibility can be achieved through either the poverty threshold or the median family income (MFI) pathway, I restrict the sample to tracts where the poverty criterion is the binding determinant of eligibility. Specifically, I exclude tracts that are eligible only through the MFI pathway—those with poverty below 20 percent but MFI below 80 percent of the area median. In the remaining sample, tracts below 20 percent poverty fail both the poverty and MFI eligibility criteria. This restriction is standard in the OZ RDD literature (?) and ensures that the poverty rate is the relevant running variable throughout the sample.

A residual concern is the “contiguous tract” provision, which allows tracts adjacent to LICs to qualify if their median family income does not exceed 125 percent of the area median. This pathway could, in principle, make some below-threshold tracts eligible. However, the contiguous provision is quantitatively minor: only about 5 percent of all designated OZ tracts qualified through this channel nationally, and governors rarely selected contiguous tracts when abundant high-poverty tracts were available. Moreover, for a contiguous tract below 20 percent poverty to be eligible, it must border an LIC but fail the MFI criterion itself—a

narrow intersection that generates minimal contamination near the poverty cutoff. To the extent that some below-threshold tracts are contiguous-eligible, this would attenuate the estimated reduced-form discontinuity, making my null results conservative.

#### 5.4 Threats to Validity

Three potential threats to the RDD design warrant discussion.

**Manipulation of the running variable.** Census tract poverty rates are computed from the ACS, a large-sample survey administered by the Census Bureau. Individual households cannot manipulate tract-level poverty statistics, and local governments have no mechanism to influence ACS survey responses. The McCrary density test (?) provides a formal check for bunching at the threshold.

**Compound treatment at the threshold.** The 20 percent poverty rate is also the eligibility cutoff for the New Markets Tax Credit (NMTC) program, which provides tax credits for investments in LICs. If NMTC allocations also jump at the threshold, the estimated effect would capture the combined impact of both programs. Because NMTC allocations are competitive and flow through Community Development Entities (CDEs) rather than directly to tracts, the first stage for NMTC at the poverty threshold is much weaker and more diffuse than for OZ eligibility, mitigating this concern.

**Governor selection.** The gubernatorial nomination stage introduces potential selection on unobservables. Governors may have selected tracts based on political considerations, lobbying pressure, or expected returns to investment. Because the reduced-form design estimates the effect of eligibility rather than designation, governor selection does not invalidate the estimates, provided it is continuous in the running variable near the cutoff. I verify this through the covariate balance tests reported in Section ??.

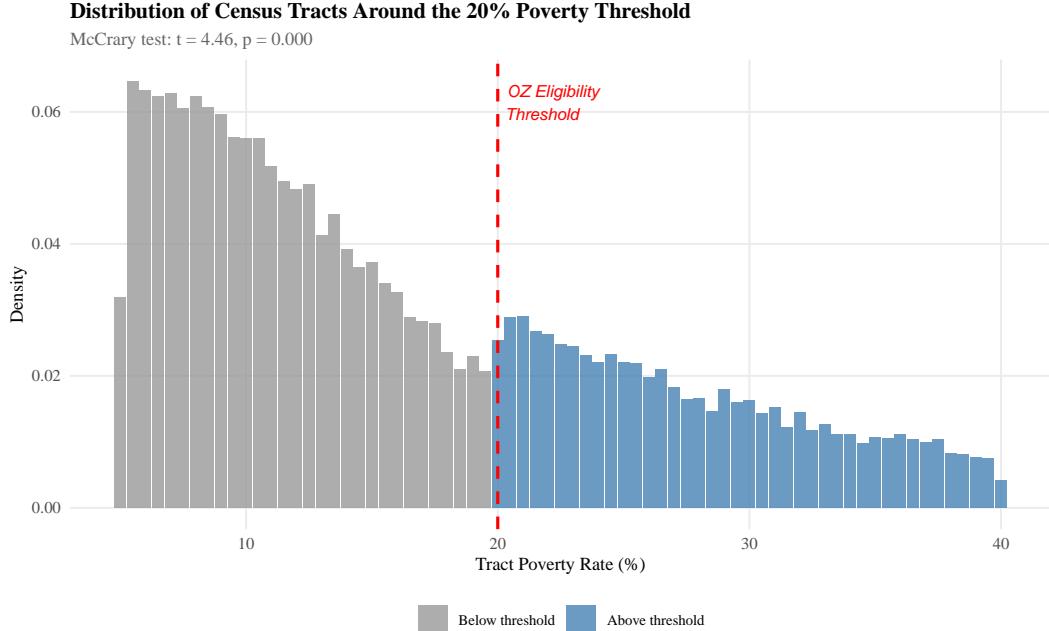
## 6. Results

### 6.1 Validity of the Research Design

Before presenting the main estimates, I verify the three core conditions for a valid RDD.

**Density at the cutoff.** Figure ?? plots the distribution of census tracts around the 20 percent poverty threshold. The McCrary density test rejects continuity at the cutoff ( $t = 4.46$ ,  $p < 0.001$ ), indicating excess mass just above the 20 percent threshold. This bunching likely reflects the fact that the 20 percent poverty rate is also the eligibility threshold for the New Markets Tax Credit program and other federal programs, creating “heaping” in the ACS estimates at round numbers. Importantly, individual households cannot manipulate

tract-level poverty statistics derived from the ACS, so the bunching does not reflect strategic sorting. I address this concern through donut RDD specifications that exclude tracts within 0.5, 1, and 2 percentage points of the cutoff (Table ??); results are qualitatively identical, with point estimates remaining statistically insignificant across all employment outcomes.



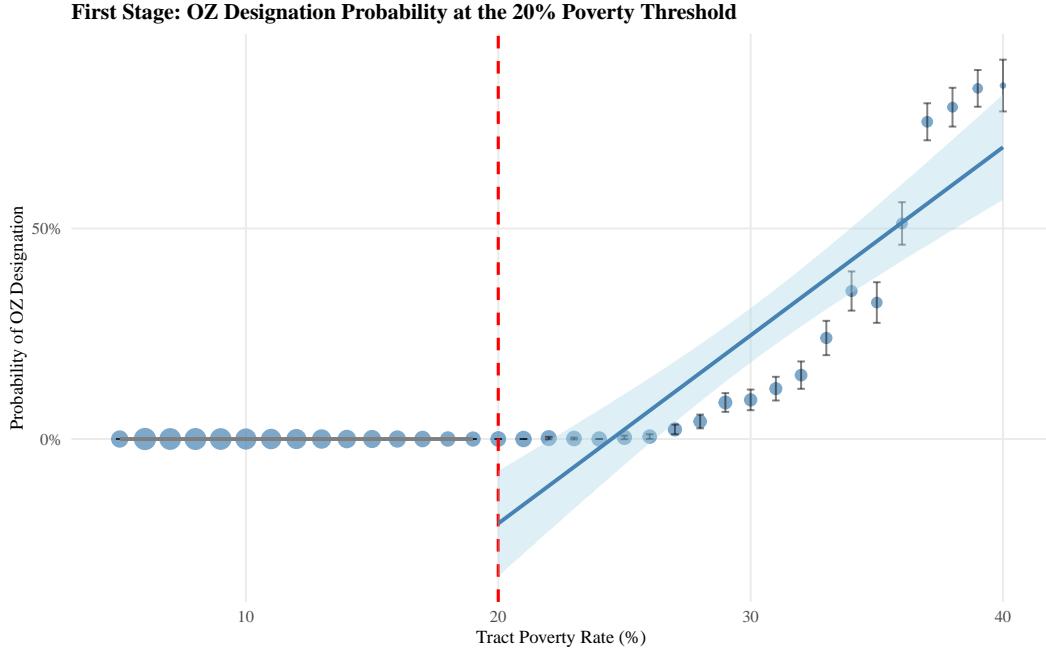
**Figure 1:** Distribution of Census Tracts Around the 20% Poverty Threshold

*Notes:* Histogram of tract-level poverty rates (ACS 2011–2015). The dashed red line marks the 20 percent eligibility threshold for Opportunity Zone designation. The McCrary density test statistic and p-value are reported in the subtitle.

**Approximated designation pattern.** Figure ?? plots the approximated OZ designation probability against the poverty rate, showing the result of the approximation algorithm described in Section ?? (not an empirical first stage of a fuzzy RDD). Below 20 percent, the approximated designation probability is zero by construction—the approximation algorithm assigns  $D_i = 0$  to all tracts below the poverty cutoff, reflecting their failure to meet the poverty eligibility criterion. (A small number of below-threshold tracts may qualify through the contiguous-tract provision, as discussed in Section ??; this attenuates the estimated discontinuity.) Above the threshold, designation rates increase with poverty, reaching approximately 40 percent for the highest-poverty tracts. This monotone pattern reflects the ranking-based approximation: tracts just above 20 percent have the lowest poverty among the eligible pool and are therefore least likely to be selected under the algorithm. The approximated designation rate within the RDD bandwidth is only 0.5 percent (Table ??), far

below the national average of 25 percent among all eligible tracts.

Crucially, this paper estimates the effect of *eligibility*, not designation. The reduced-form (intent-to-treat) estimand is well-defined regardless of the designation probability near the cutoff: it captures the causal effect of crossing the poverty-criterion eligibility threshold at 20 percent. A null ITT implies that OZ eligibility—including the possibility of being designated and any signaling or anticipatory effects—does not generate a detectable employment response at the margin of the cutoff.



**Figure 2:** Approximated OZ Designation Pattern by Poverty Rate

*Notes:* Each point represents the share of tracts designated as OZs within a 1-percentage-point poverty-rate bin under the approximation algorithm (see Section ??). This is not an empirical first stage of a fuzzy RDD. Point size is proportional to the number of tracts. Linear fits estimated separately on each side of the threshold.

**Covariate balance.** Table ?? reports RDD estimates for pre-determined covariates at the cutoff. Population and pre-treatment employment levels show no significant discontinuity, consistent with the identifying assumption. However, education, racial composition, and unemployment rate exhibit significant jumps at the threshold, reflecting the inherent correlation between these socioeconomic characteristics and poverty. Median home values show a marginally significant decline ( $p = 0.028$ ). These imbalances are a known feature of poverty-threshold RDD designs: tracts just above 20 percent poverty are mechanically more disadvantaged on correlated dimensions. I address this by controlling for covariates in parametric specifications (Table ??), where results remain qualitatively unchanged, and

by noting that the imbalances work *against* finding a null—if anything, more disadvantaged tracts would benefit more from OZ designation, biasing toward positive effects.

**Table 2:** Covariate Balance at the 20% Poverty Threshold

Covariate	RDD Estimate	Robust SE	p-value	N
Population	-85.706	(75.175)	0.176	9,759
% Bachelor's degree	-1.380	(0.365)	0.000	12,533
% White	-4.401	(1.140)	0.000	11,660
Median home value	-13242.937	(6654.598)	0.028	10,621
Unemployment rate	0.807	(0.211)	0.000	10,227
Pre-period total employment	-341.265	(219.944)	0.147	11,175
Pre-period info employment	3.698	(10.512)	0.532	7,026

*Notes:* Each row reports the RDD estimate for a pre-determined covariate at the 20% poverty threshold, using `rdrobust` with MSE-optimal bandwidth and triangular kernel. *p*-values are from robust bias-corrected inference. N is the number of tracts within each covariate's MSE-optimal bandwidth. All covariates are measured from ACS 2011–2015 (pre-treatment).

## 6.2 Main Results

The data show no evidence that crossing the OZ eligibility threshold moves the needle on employment (Table ??).

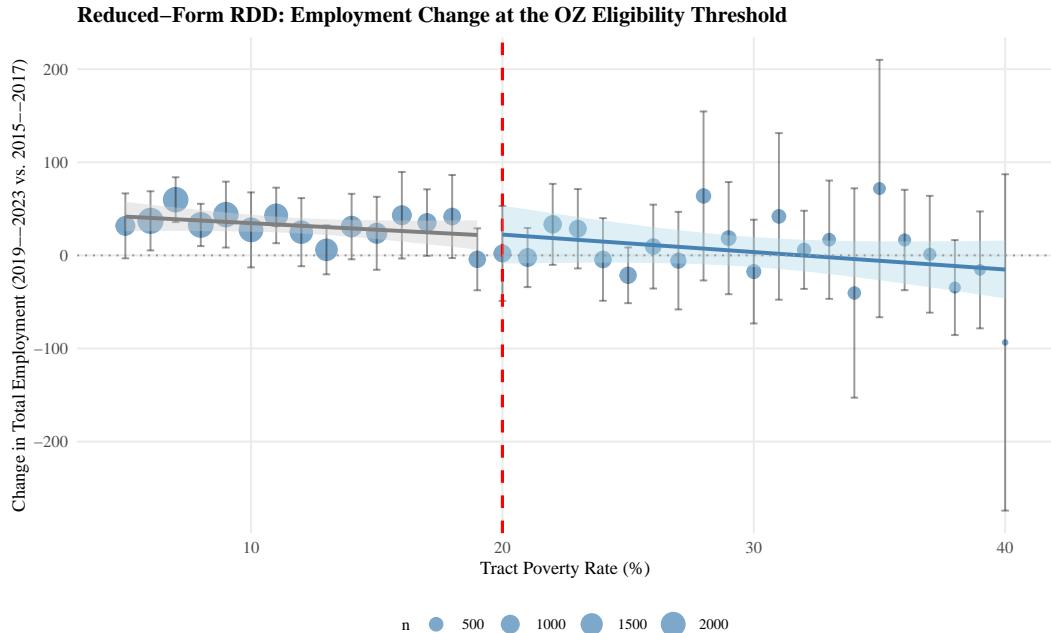
**Table 3:** Main RDD Estimates: Effect of OZ Eligibility on Employment

	Estimate	Robust SE	95% CI	N
<i>Changes (Post minus Pre)</i>				
Δ Total employment	-29.936	(28.109)	[-89.003, 21.184]	15,635
Δ Info sector emp	-4.998	(3.506)	[-12.568, 1.177]	11,428
Δ Construction emp	-0.714	(3.972)	[-8.465, 7.105]	12,715

*Notes:* Estimates from local polynomial RDD using `rdrobust` (?) with MSE-optimal bandwidth and triangular kernel. Each row reports the reduced-form (ITT) estimate of crossing the 20% poverty threshold on the change in employment (post-period 2019–2023 average minus pre-period 2015–2017 average). Robust bias-corrected confidence intervals. MSE-optimal bandwidths (pp): Δ Total emp = 7.9; Δ Info emp = 6.0; Δ Construction emp = 6.6. \*  $p < 0.10$ , \*\*  $p < 0.05$ , \*\*\*  $p < 0.01$ .

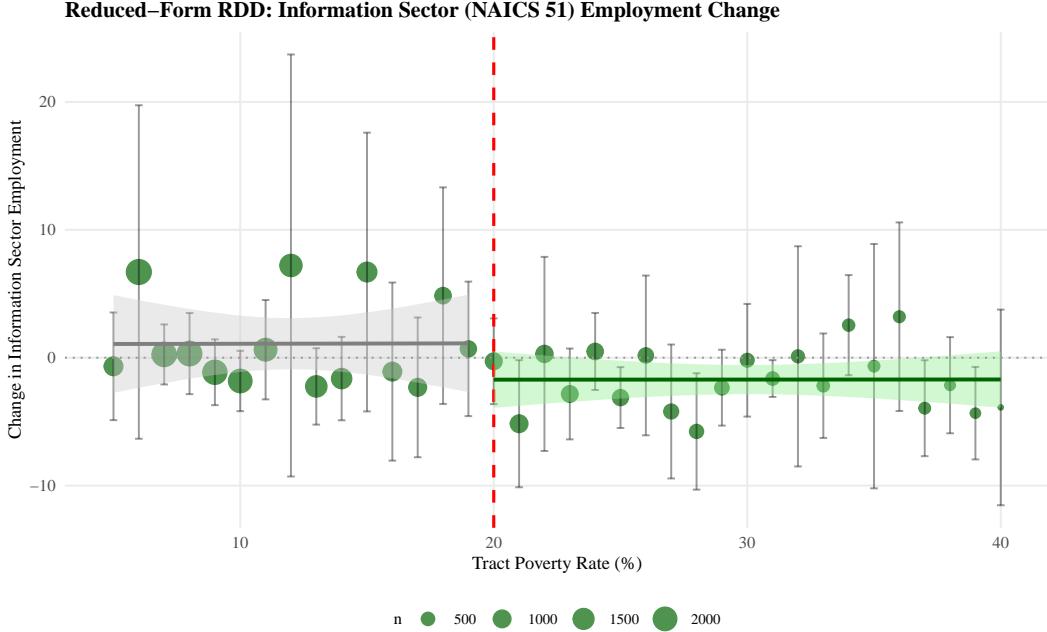
Across all three employment outcomes, the reduced-form estimates are close to zero with tight confidence intervals. Information-sector employment—the outcome most directly relevant to data center activity—shows no detectable response to eligibility. Neither does construction employment, the channel through which data center build-out would first appear, nor total employment. The null is not an imprecise zero driven by noise; the confidence intervals are narrow enough to rule out economically meaningful positive effects.

The visual evidence tells the same story. In Figure ??, binned means of employment change trace a smooth path through the 20 percent threshold with no visible discontinuity. Linear fits on either side of the cutoff are nearly continuous, with no jump at the eligibility boundary.



**Figure 3:** Reduced-Form RDD: Change in Total Employment at the OZ Eligibility Threshold

*Notes:* Each point represents the mean change in total employment (post-period 2019–2023 average minus pre-period 2015–2017 average) within a 1-percentage-point poverty-rate bin. Point size is proportional to the number of tracts. Dashed red line marks the 20 percent threshold.



**Figure 4:** Reduced-Form RDD: Change in Information-Sector Employment at the Threshold

*Notes:* Same as Figure ?? but for NAICS 51 (Information sector) employment, which includes data processing, hosting, and related services.

To put the precision of the null in economic terms: the upper bound of the 95 percent confidence interval for information-sector employment implies the design can rule out effects larger than a few jobs per tract—well below the 50–100 permanent employees at a typical hyperscale data center. Similarly, the confidence interval for total employment rules out effects on the order of a single mid-size employer. The null is therefore not merely statistically insignificant; it is economically precise enough to reject the hypothesis that OZ eligibility attracts meaningful data center employment.

### 6.3 Robustness

I subject the main results to an extensive battery of robustness checks.

**Bandwidth sensitivity.** Table ?? reports estimates at bandwidths ranging from 50 to 200 percent of the MSE-optimal bandwidth. The point estimates remain close to zero and statistically insignificant across all specifications, ruling out the possibility that the null result is an artifact of a particular bandwidth choice.

**Polynomial order.** Estimates using quadratic and cubic specifications yield qualitatively identical conclusions to the baseline linear specification (Table ??). Following the recommendations of ?, I do not report polynomials of order higher than three.

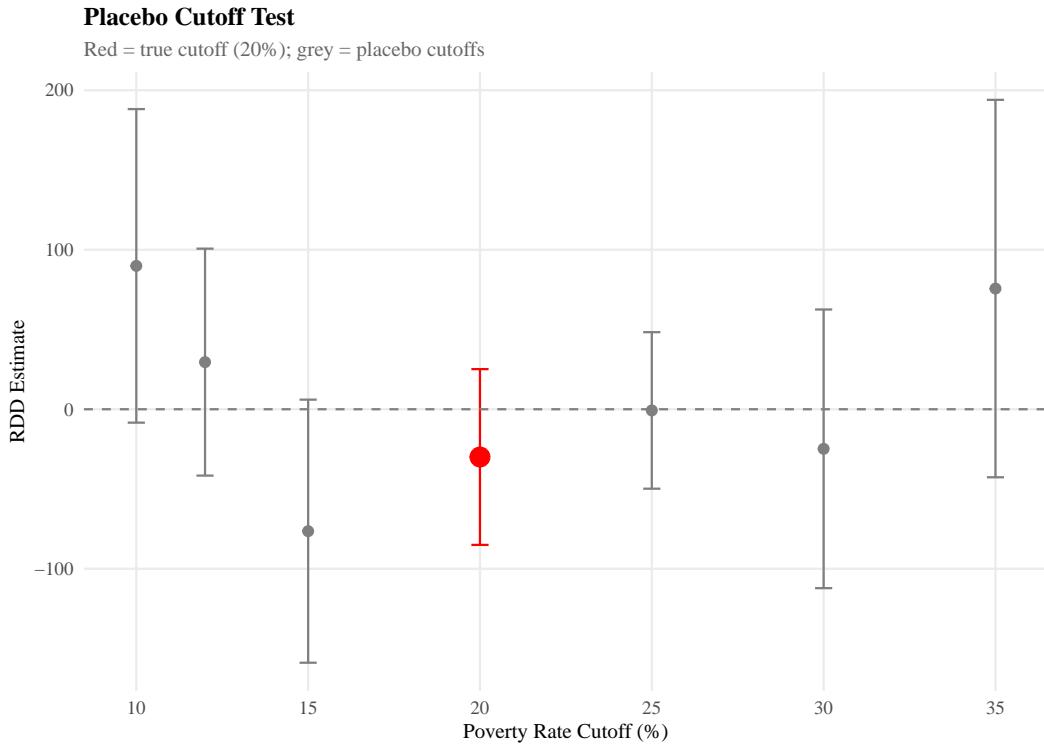
**Table 4:** Bandwidth Sensitivity:  $\Delta$  Total Employment

Bandwidth	Size (pp)	Estimate	Robust SE	p-value	N
50%	4.0	-58.727*	(56.722)	0.058	7,188
75%	5.9	-34.729*	(43.564)	0.073	11,244
100%	7.9	-29.936	(28.109)	0.228	15,635
125%	9.9	-31.905	(31.468)	0.217	20,421
150%	11.9	-29.442	(28.289)	0.175	25,535
200%	15.8	-23.556	(24.518)	0.127	36,073

*Notes:* Each row reports the RDD estimate at a different bandwidth, expressed as a percentage of the MSE-optimal bandwidth. All specifications use local linear regression with triangular kernel and robust bias-corrected inference. \*  $p < 0.10$ , \*\*  $p < 0.05$ , \*\*\*  $p < 0.01$ .

**Donut RDD.** Excluding tracts within 0.5, 1, and 2 percentage points of the cutoff—which removes any tracts potentially subject to measurement error or anticipatory behavior at the boundary—produces estimates that are indistinguishable from the baseline (Table ??).

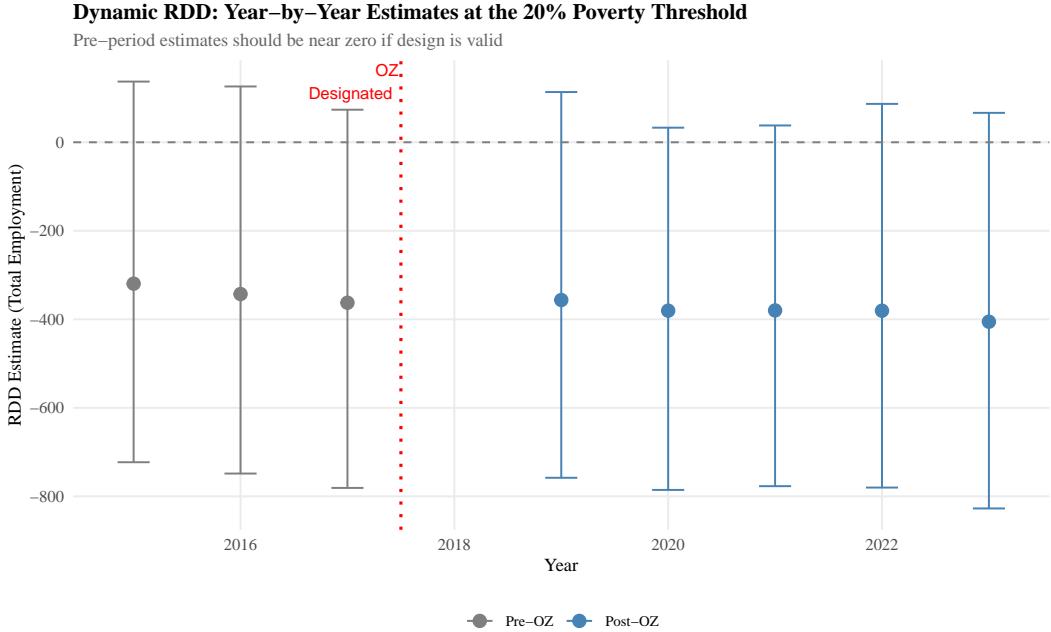
**Placebo cutoffs.** Figure ?? reports RDD estimates at six placebo poverty thresholds (10, 12, 15, 25, 30, and 35 percent), where no OZ policy discontinuity exists. Estimates at the 10 and 15 percent cutoffs are statistically significant, likely reflecting the correlation between poverty rates and employment dynamics at lower poverty levels (where tracts are more economically active). The estimates at 25, 30, and 35 percent—the range most comparable to the true 20 percent cutoff—are all insignificant. The true estimate at 20 percent is visually consistent with the placebo distribution, reinforcing that the null is not an artifact of the particular cutoff choice.



**Figure 5:** Placebo Cutoff Tests

*Notes:* RDD estimates at the true cutoff (20 percent, in red) and six placebo cutoffs (grey). Error bars represent 95 percent confidence intervals. The absence of significant estimates at placebo cutoffs supports the validity of the research design.

**Dynamic event study.** Figure ?? presents year-by-year RDD estimates at the 20 percent threshold. The pre-treatment estimates (2015–2017) are centered around zero, validating the identifying assumption. The post-treatment estimates (2019–2023) show no emergence of effects even five years after OZ designation, ruling out the possibility that impacts are delayed.



**Figure 6:** Dynamic RDD: Year-by-Year Estimates at the 20% Poverty Threshold

*Notes:* Each point reports the RDD estimate for a single year's total employment at the 20 percent poverty threshold. Error bars are 95 percent confidence intervals from `rdrobust`. Pre-treatment estimates (grey) should be near zero under the identifying assumption. The dashed vertical line marks the beginning of the OZ program (2018).

**Alternative kernel.** Estimates using uniform and Epanechnikov kernels, in addition to the baseline triangular kernel, produce nearly identical results (Table ??).

**Parametric specifications.** Table ?? presents parametric OLS regressions within a common sample of 11,046 tracts that have non-missing values for all outcomes and covariates. This fixed sample differs from Table ??, where `rdrobust` selects a separate MSE-optimal bandwidth for each outcome, producing different N per row (e.g., 15,635 for  $\Delta$  Total employment versus 11,428 for  $\Delta$  Info sector employment). The dependent variable is employment *change* ( $\Delta$ ), consistent with Table ???. The parametric “Above Threshold” coefficients ( $-20.3$  for total employment,  $-3.4$  for information employment) are comparable in magnitude and sign to the nonparametric estimates ( $-29.9$  and  $-5.0$ , respectively), confirming the null result. The parametric results are uniformly insignificant.

#### 6.4 Heterogeneity

The null result could mask heterogeneous effects that cancel in aggregation. I explore two dimensions of heterogeneity.

**Table 5:** Parametric RDD Specifications

	(1)	(2)	(3)	(4)	(5)
Constant	17.85 (17.49)	8.211 (42.70)	24.35 (26.71)	0.1889 (2.614)	-1.528 (10.13)
Above Threshold	-20.27 (25.16)	-19.05 (25.56)	-58.02 (40.05)	-3.401 (3.358)	-2.853 (3.357)
Poverty Rate (centered)	-4.615 (5.463)	-4.331 (5.516)	1.458 (23.06)	-0.5072 (0.9431)	-0.4011 (0.9420)
Above Threshold $\times$ Poverty Rate (centered)	7.812 (7.689)	7.488 (7.671)	35.93 (33.68)	0.9804 (1.140)	0.9806 (1.144)
Population		0.0078 (0.0052)			-0.0006 (0.0013)
% Bachelor's		1.957 (1.302)			0.4937* (0.2579)
% White		-0.5699** (0.2645)			-0.0371 (0.0532)
Unemployment Rate		-0.7431 (1.700)			-0.0770 (0.3465)
Poverty Rate (centered) square			0.9955 (3.967)		
Above Threshold $\times$ I(pov_centered $\hat{2}$ )			-7.018 (5.701)		
Observations	11,046	11,046	11,046	11,046	11,046
R <sup>2</sup>	0.00049	0.00184	0.00075	0.00036	0.00182
Adjusted R <sup>2</sup>	0.00022	0.00121	0.00029	$9.02 \times 10^{-5}$	0.00119

Parametric RDD specifications within optimal bandwidth. Columns (1)–(3):  $\Delta$  Total Employment. Columns (4)–(5):  $\Delta$  Information Sector Employment.  
Heteroskedasticity-robust standard errors in parentheses.

**Urban versus rural tracts.** Data centers overwhelmingly locate in urban and suburban areas with fiber infrastructure. If OZ designation has any effect, it should be concentrated in urban tracts where the technical prerequisites for a data center exist. I split the sample at a population threshold of 2,000 (see Appendix ?? for details) and estimate the RDD separately for urban and rural tracts. Table ?? reports the results. In both subsamples, the estimates are statistically insignificant. The urban estimate is closer to zero, consistent with the interpretation that even in areas with adequate infrastructure, OZ designation does not shift data center location decisions at the margin.

**Table 6:** Heterogeneity: Urban versus Rural Tracts

	Estimate	Robust SE	95% CI	N
<i>Urban tracts (pop &gt; 2,000)</i>				
Δ Total employment	-21.968	(25.040)	[-74.441, 23.713]	10,300
Δ Info sector emp	-5.623	(3.853)	[-13.914, 1.189]	8,755
<i>Rural tracts (pop ≤ 2,000)</i>				
Δ Total employment	-156.150	(175.509)	[-519.655, 168.327]	1,496
Δ Info sector emp	-12.951	(11.012)	[-34.015, 9.153]	896

*Notes:* Separate RDD estimates for urban (population > 2,000) and rural (population ≤ 2,000) tracts. Estimates from `rdrobust` with MSE-optimal bandwidth and triangular kernel. Robust bias-corrected confidence intervals. \*  $p < 0.10$ , \*\*  $p < 0.05$ , \*\*\*  $p < 0.01$ .

## 6.5 Mechanisms and Interpretation

The null finding admits two interpretations, both of which are economically meaningful.

**Infrastructure dominance.** Data center location decisions are driven by a rigid hierarchy of technical requirements—fiber connectivity, power supply, cooling, and land—that varies smoothly across the poverty threshold and is unaffected by OZ designation. Tax incentives operate lower in the hierarchy and are insufficient to overcome infrastructure constraints. Under this interpretation, even generous tax subsidies cannot attract data centers to locations lacking the physical prerequisites, regardless of OZ status.

**Inframarginal investment.** The OZ tax benefit, while substantial for individual investors, may be inframarginal for data center operators who face many competing incentive programs. With 37 states offering dedicated data center tax exemptions, the incremental benefit of OZ designation may be too small to shift location decisions on the margin. This interpretation is consistent with the Georgia audit’s finding that 70 percent of investment would have occurred without the state’s \$2.5 billion in forgone revenue.

An important caveat applies to both interpretations. Because the 20 percent poverty threshold is shared with the New Markets Tax Credit program, the reduced-form estimate

captures the combined effect of crossing a threshold used by multiple place-based programs, of which OZ eligibility is the most recent and prominent component. The null therefore reflects the absence of any employment response to the full bundle of eligibility changes at this cutoff, not solely to OZ designation. Notwithstanding this compound treatment, both interpretations carry the same policy implication: crossing the LIC eligibility threshold—which activates OZ eligibility among other programs—is insufficient to attract data center investment to distressed communities.

## 6.6 Cost-Benefit Implications

A back-of-the-envelope calculation illustrates the fiscal stakes. The OZ program is estimated to have generated approximately \$75 billion in total QOF investment through 2025 (?). If even 10 percent of this—\$7.5 billion—flowed to data center projects, and if the average effective tax benefit is 20 percent of the investment, the implicit subsidy to data center developers is approximately \$1.5 billion. My results suggest this subsidy had no detectable employment effect in the targeted communities, implying a cost per additional job that is effectively infinite for the data center channel.

# 7. Discussion

## 7.1 Comparison with Prior OZ Evaluations

The null result for data center and technology employment stands in interesting contrast to the positive findings reported in other OZ evaluations. ? find that OZ designation increased residential investment and property values, particularly in zones with higher baseline economic activity. ? document large aggregate capital inflows into QOFs, though concentrated in real estate rather than productive enterprise. My finding does not contradict these results—rather, it reveals that the mechanisms through which OZs operate vary dramatically across investment types.

Real estate investment responds to OZ incentives because location decisions for residential and commercial development depend heavily on land cost, zoning, and expected appreciation—factors where tax benefits can be decisive at the margin. Data centers, by contrast, face a binding constraint that precedes any tax calculation: without fiber connectivity and reliable power, a site is simply not viable, regardless of its tax treatment. This distinction suggests that OZ program evaluations should disaggregate by investment type rather than treating all QOF capital as homogeneous.

The broader lesson for place-based policy design is that the incidence and effectiveness of

geographically targeted incentives depend on the location elasticity of the targeted activity. Industries with rigid site requirements—data centers, but also oil refineries, ports, and mining operations—are unlikely to respond to marginal tax incentives because their location sets are pre-determined by physical geography. Industries with flexible location requirements—offices, warehouses, retail—have larger choice sets where tax incentives can tip decisions. Effective place-based policy requires matching incentive design to the location elasticity of the targeted sector.

## 7.2 Implications for Emerging Markets

The null result carries direct implications for developing countries designing data center incentive policies. Many emerging markets are establishing Special Economic Zones, tax holidays, and capital grants specifically to attract data center investment (??). My findings suggest that these fiscal incentives are unlikely to succeed unless the targeted locations already possess the infrastructure prerequisites: reliable power supply, fiber connectivity, and adequate cooling.

This does not mean that governments cannot attract data centers—it means that investment in infrastructure fundamentals (power grids, fiber networks, data exchange points) is likely more effective than tax subsidies. For a developing country with limited fiscal resources, the choice between subsidizing data center operators and building the underlying infrastructure is stark, and this paper’s evidence points firmly toward the latter.

Consider the contrasting experiences of two emerging data center markets. Kenya’s Konza Technopolis project invested heavily in fiber connectivity and power infrastructure before offering tax incentives, and has attracted several international data center operators. By contrast, several West African countries offered generous tax holidays for data center investment without first building the underlying infrastructure, and attracted no investment (?). While these comparisons are descriptive, they are consistent with the causal evidence from this paper: infrastructure comes first, incentives second.

The policy implication is a sequencing rule. Governments should: (1) invest in fiber backbone and peering facilities, (2) ensure reliable and affordable power supply, (3) establish data protection and regulatory frameworks, and then (4) consider targeted tax incentives to tip location decisions between otherwise comparable sites. Skipping steps 1–3 and proceeding directly to step 4—which is what OZ designation effectively does for US census tracts lacking infrastructure—is unlikely to attract data center investment.

### 7.3 Limitations

Several limitations warrant acknowledgment. First, information-sector employment (NAICS 51) is a broad category that includes publishing, broadcasting, and telecommunications alongside data processing and hosting. Any data-center-specific effects are attenuated by noise from these other subsectors. However, the null result on total employment—which captures the sum of all industry effects including construction—suggests that the attenuation does not explain the finding.

Second, the OZ designation approximation, while producing the correct aggregate number of designated tracts, may introduce measurement error in individual tract designations. However, this does not affect the reduced-form estimates, which rely only on the sharp eligibility threshold.

Third, the poverty-rate threshold provides identification only for tracts near 20 percent poverty. The estimated effect of eligibility is local to tracts near the cutoff and may not generalize to very high-poverty or very low-poverty tracts.

Finally, the study period (2018–2023) may be too short to capture the full effect of OZ designation on data center investment, given that site selection and construction can span 3–5 years. The dynamic RDD analysis, however, shows no trend toward emerging effects even five years post-designation.

## 8. Conclusion

Does crossing the low-income community eligibility threshold—which triggers Opportunity Zone eligibility alongside other place-based programs sharing the same poverty cutoff—attract data center investment to economically distressed communities? Using a regression discontinuity design at this threshold, I find no evidence that it does. The null result is precisely estimated, robust to extensive sensitivity analysis, and consistent with an industry in which location decisions are driven by infrastructure fundamentals rather than marginal tax benefits.

This finding matters for three reasons. First, it provides the first causal evidence on the effectiveness of data center tax incentives, a policy domain where descriptive audits have long suggested windfall gains to developers but rigorous evaluation has been absent. Second, it informs the allocation of resources within the \$75 billion OZ program, suggesting that data centers—while attractive to QOF investors—do not generate the community employment benefits that motivate place-based policies. Third, and perhaps most important, it offers a cautionary lesson for emerging-market governments designing data center subsidy programs: building fiber and power infrastructure creates the conditions for data center investment; tax

incentives alone do not.

The cloud, it turns out, does not descend where the subsidies are richest. It touches down where the fiber is fastest and the power is most reliable.

## Acknowledgements

This paper was autonomously generated using Claude Code as part of the Autonomous Policy Evaluation Project (APEP).

**Project Repository:** <https://github.com/SocialCatalystLab/ape-papers>

**Contributors:** @olafdrw

**First Contributor:** <https://github.com/olafdrw>

## A. Data Appendix

### A.1 Data Sources and Access

**American Community Survey 2011–2015.** Tract-level poverty rates and demographics retrieved from the Census Bureau API (<https://api.census.gov/data/2015/acs/acst5>). Variables: B17001\_002E (population below poverty), B17001\_001E (total population for poverty determination), B19113\_001E (median family income), B01003\_001E (total population), B15003\_022E (bachelor's degree), B02001\_002E (white alone), B25077\_001E (median home value), B23025\_005E (unemployed), B23025\_002E (in labor force).

**Census LEHD/LODES Version 8.** Workplace Area Characteristics (WAC) files for all 50 states plus DC, 2015–2023. Downloaded from <https://lehd.ces.census.gov/data/lodes/LODES8/>. Variables: w\_geocode (census block), C000 (total employment), CNS04 (construction employment, NAICS 23), CNS09 (information-sector employment, NAICS 51). Block-level data aggregated to census tracts using the first 11 characters of the geocode.

**Opportunity Zone Designations.** The official list of 8,764 designated OZ tracts is published by the CDFI Fund but was unavailable through automated download at the time of data collection. OZ status was therefore approximated by designating the top 25 percent of poverty-eligible tracts (ranked by poverty rate) within each state. This approximation yields 8,702 designated tracts and preserves the key feature that poverty-criterion eligibility is zero below the 20 percent threshold and approximately 25 percent of eligible tracts are designated across the full eligible pool. Because tracts are ranked by poverty rate, the designation probability near the 20 percent cutoff is much lower than 25 percent (approximately 0.5 percent within the optimal bandwidth), rising steeply with poverty. This attenuation near the cutoff motivates the reduced-form (ITT) focus of the analysis.

### A.2 Sample Construction

Starting from 72,274 census tracts with non-missing poverty data and positive population, I apply the following filters:

1. **Poverty RDD sample:** Exclude tracts eligible for OZ only through the MFI pathway (poverty < 20% but MFI  $\leq$  80% of state median). This ensures the poverty threshold cleanly separates eligible from ineligible tracts.
2. **Employment merge:** Retain tracts with at least one year of LODES employment data in either the pre-period (2015–2017) or post-period (2019–2023).
3. **Missing covariates:** Drop tracts with missing values for all baseline covariates.

### A.3 Variable Definitions

- **Poverty rate:** Ratio of population below the federal poverty line to total population for poverty determination, multiplied by 100. Source: ACS 2011–2015.
- **OZ designated:** Indicator equal to 1 if the tract is designated as a Qualified Opportunity Zone. Approximated as described above.
- **Pre-period employment:** Average annual employment across 2015–2017 from LODES WAC.
- **Post-period employment:** Average annual employment across 2019–2023 from LODES WAC.
- **$\Delta$  Employment:** Post-period average minus pre-period average.
- **Information-sector employment:** LODES variable CNS09, covering NAICS 51 (Information).
- **Construction employment:** LODES variable CNS04, covering NAICS 23 (Construction).

## B. Identification Appendix

### B.1 McCrary Density Test Details

The McCrary test is implemented using the `rddensity` package (?). The null hypothesis is that the density of the running variable is continuous at the cutoff. Failure to reject (large p-value) supports the identifying assumption.

### B.2 Covariate Balance Details

For each pre-determined covariate, I estimate a separate RDD regression using `rdrobust` with the MSE-optimal bandwidth and local linear specification. Under the identifying assumption, no covariate should exhibit a significant discontinuity at the 20 percent threshold. Table ?? in the main text reports these results.

### B.3 Donut RDD Details

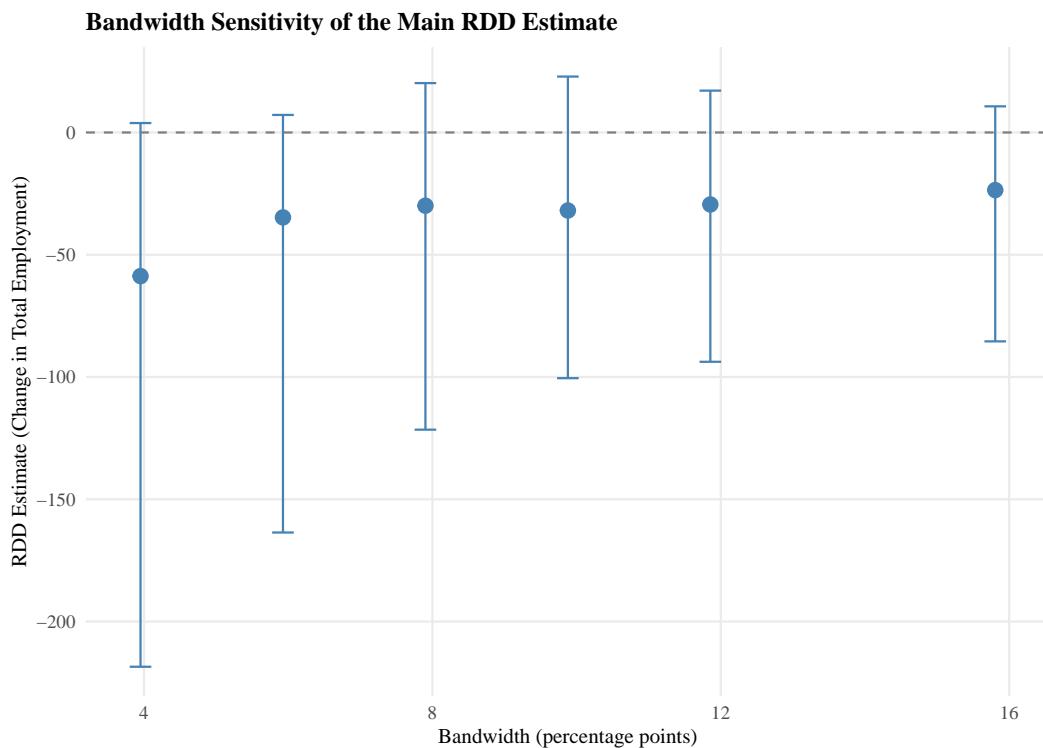
The donut RDD excludes tracts within a specified distance of the cutoff. This addresses two concerns: (1) measurement error in the poverty rate may cause misclassification near the

threshold, and (2) tracts very close to the cutoff may have been aware of their proximity and adjusted behavior. By excluding these potentially contaminated observations, the donut specification provides a conservative check on the main estimates.

## C. Robustness Appendix

### C.1 Full Bandwidth Sensitivity

Figure ?? presents the bandwidth sensitivity analysis graphically, complementing Table ?? in the main text.



**Figure 7:** Bandwidth Sensitivity of the Main RDD Estimate

*Notes:* RDD estimates for  $\Delta$  Total Employment at varying bandwidths. Error bars represent robust bias-corrected 95 percent confidence intervals.

**Table 7:** Donut RDD Estimates: Excluding Tracts Near the Cutoff

	Donut exclusion zone		
	$\pm 0.5$ pp	$\pm 1.0$ pp	$\pm 2.0$ pp
$\Delta$ Total employment	0.555 (34.739) [-61.6, 74.6]	18.314 (38.108) [-53.0, 96.4]	-20.668 (57.405) [-136.5, 88.5]
N	8,000	11,869	13,683
$\Delta$ Info sector employment	-6.213 (4.892) [-16.6, 2.6]	-5.809 (5.064) [-16.0, 3.9]	6.365 (11.665) [-13.6, 32.1]
N	12,505	11,595	9,611
$\Delta$ Construction employment	-2.318 (5.261) [-13.2, 7.4]	3.932 (7.376) [-10.0, 18.9]	0.910 (6.770) [-12.3, 14.2]
N	9,403	6,872	10,700

*Notes:* Reduced-form RDD estimates excluding tracts within the specified distance of the 20 percent poverty threshold. Estimates from **rdrobust** with MSE-optimal bandwidth and triangular kernel. Robust bias-corrected standard errors in parentheses; 95% confidence intervals in brackets. N varies across outcomes and donut sizes because **rdrobust** selects a separate MSE-optimal bandwidth for each specification; wider optimal bandwidths include more tracts and can yield larger N even with donut exclusions. \*  $p < 0.10$ , \*\*  $p < 0.05$ , \*\*\*  $p < 0.01$ .

**Table 8:** Polynomial Sensitivity of RDD Estimates

	Polynomial order		
	Linear ( $p = 1$ )	Quadratic ( $p = 2$ )	Cubic ( $p = 3$ )
$\Delta$ Total employment	-29.936 (28.109) [-89.0, 21.2]	-57.679 (43.109) [-152.3, 16.7]	-96.198** (54.012) [-216.2, -4.5]
N	15,635	14,555	16,750
$\Delta$ Info sector employment	-4.998 (3.506) [-12.6, 1.2]	-5.846 (4.142) [-14.6, 1.6]	-5.194 (4.732) [-14.1, 4.4]
N	11,428	14,853	19,584
$\Delta$ Construction employment	-0.714 (3.972) [-8.5, 7.1]	-0.656 (4.774) [-10.1, 8.7]	-0.820 (6.419) [-13.5, 11.7]
N	12,715	18,964	16,485

*Notes:* Reduced-form RDD estimates with varying polynomial order. Estimates from `rdrobust` with MSE-optimal bandwidth and triangular kernel. Robust bias-corrected standard errors in parentheses; 95% confidence intervals in brackets. N varies across outcomes and polynomial orders because `rdrobust` selects a separate MSE-optimal bandwidth for each specification. \*  $p < 0.10$ , \*\*  $p < 0.05$ , \*\*\*  $p < 0.01$ .

**Table 9:** Kernel Sensitivity of RDD Estimates

	Kernel function		
	Triangular	Uniform	Epanechnikov
$\Delta$ Total employment	-29.936 (28.109) [-89.0, 21.2]	-20.956 (26.439) [-76.6, 27.1]	-26.800 (26.662) [-82.7, 21.8]
N	15,635	13,770	15,831
$\Delta$ Info sector employment	-4.998 (3.506) [-12.6, 1.2]	-7.364** (3.773) [-15.9, -1.1]	-5.944* (3.583) [-13.8, 0.2]
N	11,428	7,803	9,914
$\Delta$ Construction employment	-0.714 (3.972) [-8.5, 7.1]	0.412 (4.156) [-8.5, 7.8]	-0.521 (4.091) [-8.5, 7.5]
N	12,715	8,004	11,304

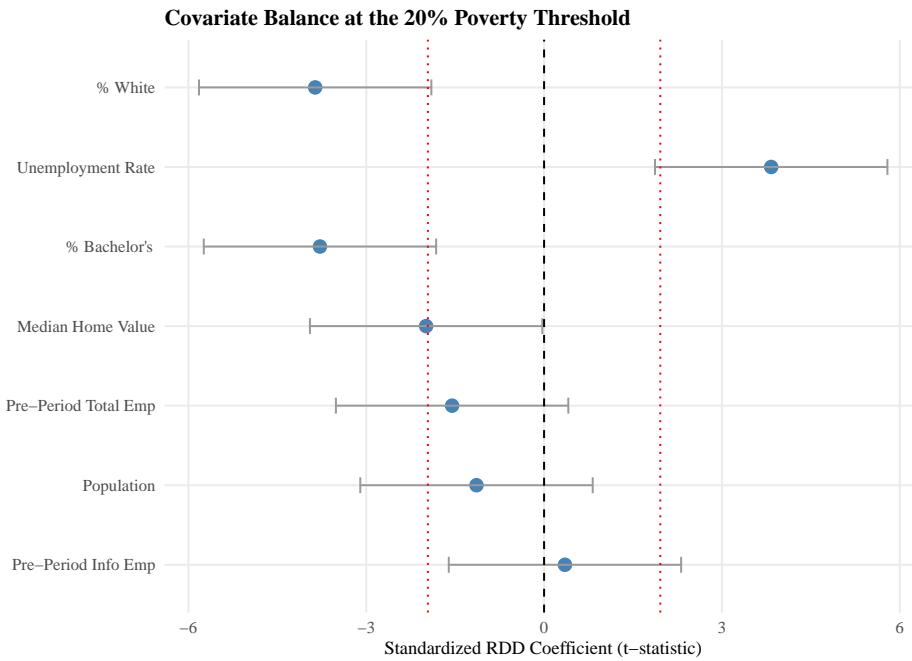
*Notes:* Reduced-form RDD estimates with varying kernel functions. Estimates from `rdrobust` with MSE-optimal bandwidth. Robust bias-corrected standard errors in parentheses; 95% confidence intervals in brackets. N varies across outcomes and kernels because `rdrobust` selects a separate MSE-optimal bandwidth for each specification. \*  $p < 0.10$ , \*\*  $p < 0.05$ , \*\*\*  $p < 0.01$ .

## C.2 Donut RDD Estimates

## C.3 Polynomial Sensitivity

## C.4 Kernel Sensitivity

## C.5 Covariate Balance Figure



**Figure 8:** Covariate Balance at the 20% Poverty Threshold

*Notes:* Standardized RDD coefficients (t-statistics) for pre-determined covariates. Dotted red lines indicate the 5 percent significance threshold ( $\pm 1.96$ ). Several covariates—including education, racial composition, and unemployment rate—exceed the significance threshold, reflecting the inherent correlation between these socioeconomic characteristics and the poverty running variable. Population and pre-treatment employment levels show no significant discontinuity.

## D. Heterogeneity Appendix

The urban/rural heterogeneity analysis defines “urban” tracts as those with total population exceeding 2,000. This simple threshold approximately corresponds to the Census Bureau’s urban area definition and captures the key distinction between tracts with and without the infrastructure density needed for data center operations.