Do Place-Based Tax Incentives Attract Data Center Investment?
Evidence from Opportunity Zones[*]

APEP Autonomous Research[†]   @olafdrw

February 23, 2026

Abstract

State-level tax incentives double data center construction, yet whether tract-level place-based programs achieve the same is unknown. I exploit the sharp poverty-rate threshold governing Opportunity Zone eligibility in a regression discontinuity design across approximately 46,000 census tracts. Using both direct data center location data—geocoded from EIA Form 860 and EPA facility records—and Census LEHD employment data, I find no effect of crossing the eligibility threshold on data center presence, information-sector employment, or total employment. The null is robust to bandwidth variation, donut specifications, local randomization inference, and placebo cutoffs. These results complete an emerging incentive hierarchy: state sales tax exemptions attract data centers but not jobs; local tax abatements attract facilities but raise public costs; OZ capital gains incentives attract neither facilities nor jobs to distressed tracts.

JEL Codes: H25, R11, R38, L86
Keywords: Opportunity Zones, data centers, place-based policies, regression discontinuity, tax incentives, digital infrastructure

---

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 (Carl Vinson Institute of Government, 2025). Georgia is not alone: 37 states now offer targeted data center tax incentives (Good Jobs First, 2025), and global data center investment has exceeded $300 billion annually since 2018 (Cisco Systems, 2023). 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?

Two concurrent papers sharpen this question. Gargano and Giacoletti (2025) exploit staggered adoption of state-level data center incentives and find that these programs *double* facility construction—increasing hyperscale facilities by 50 percent and UPS capacity by 200 percent—but generate no detectable gains in technology employment. Jaros (2026) documents that even when tax abatements successfully attract data centers to localities, public expenditures rise, creating fiscal costs that persist for decades. Together, these findings establish that state-level incentives can attract data center bricks and mortar, but the employment and fiscal channels through which communities would benefit are weak or negative. What remains unknown is whether *tract-level* place-based incentives—which target distressed neighborhoods rather than entire states—can attract data centers at all.

I answer this question by exploiting the Opportunity Zone (OZ) program, a capital gains tax incentive created by the 2017 Tax Cuts and Jobs Act that designated approximately 8,764 census tracts for preferential treatment. The designation process creates a natural experiment: census tracts were eligible for OZ designation only if their poverty rate exceeded 20 percent. Governors then selected roughly 25 percent of eligible tracts. My regression discontinuity design compares tracts just above and below this threshold.

The key innovation of this revision is direct measurement of data center presence. Previous versions relied on NAICS 51 (Information sector) employment as a proxy—a measure that includes publishing, broadcasting, and telecommunications alongside data processing. I now geocode actual data center locations from EIA Form 860 generator-level records and EPA Greenhouse Gas Reporting Program facility data to census tracts, creating binary and count measures of data center presence. This addresses the central criticism that NAICS 51 is too noisy to detect data-center-specific effects.

Using both direct data center location data and Census LEHD employment data for approximately 46,000 tracts, I find a precisely estimated null: crossing the OZ eligibility threshold has no detectable effect on data center presence, information-sector employment,

construction employment, or total employment. The null persists across all robustness checks: alternative bandwidths, donut RDD designs, systematic placebo cutoffs, local randomization inference for discrete running variables (Kolesár and Rothe, 2018; Frandsen, 2017), and infrastructure heterogeneity splits. The minimum detectable effect for data center presence implies the design can rule out increases in the probability of hosting a data center larger than a few percentage points—well within the range of effects found by Gargano and Giacoletti (2025) at the state level.

These results complete an emerging hierarchy of data center incentive effectiveness. At the broadest geographic scale, state-level sales and property tax exemptions attract large data center facilities but generate no technology employment gains (Gargano and Giacoletti, 2025). At the local level, property tax abatements attract data centers but raise public expenditures, creating fiscal costs that persist for 20–30 years (Jaros, 2026). At the tract level, OZ capital gains incentives do not attract data centers to distressed tracts, nor do they generate employment gains (this paper). This hierarchy transforms the paper from a null result on one policy to a contribution mapping the frontier of which incentives work and which do not. The pattern has a clear economic logic: state-level sales tax exemptions on equipment purchases worth millions per year are structurally different from OZ capital gains deferrals for investors. Data center operators respond to direct cost reductions on their primary expenditures (electricity, equipment), not to investor-level capital gains treatment that operates through an intermediary (the Qualified Opportunity Fund). The OZ incentive targets the wrong margin.

A caveat on interpretation: my estimates identify the effect of crossing the Low-Income Community (LIC) eligibility threshold at 20 percent poverty, which triggers OZ eligibility alongside other programs such as the New Markets Tax Credit. The reduced-form estimates capture the bundled effect of this threshold crossing; the fuzzy RDD recovers the LATE of OZ designation specifically for employment outcomes. Data center presence results are intent-to-treat. Throughout, I am careful to distinguish what the design identifies from what it does not.

We first describe the geocoded facility data and employment outcomes, then present the RDD evidence under both continuity-based and design-based frameworks, and conclude by situating these results within the broader landscape of data center incentives.

2. Related Literature

This paper connects four strands of the economics literature: place-based policies, data center tax incentives specifically, the economics of infrastructure investment, and Opportunity Zone

evaluations.

2.1 Place-Based Policies

The theoretical case for place-based policies rests on agglomeration economies, labor market frictions, and equity considerations (Bartik, 1991; Kline and Moretti, 2013; Gaubert et al., 2021). Glaeser (2008) 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 has produced mixed results. Busso et al. (2013) find that the federal Empowerment Zone program generated wage increases of 8–13 percent for zone workers, with follow-up work confirming the importance of estimator choice for these conclusions (Busso et al., 2014). In contrast, Neumark and Simpson (2015) conclude in their comprehensive review that many state and local programs fail cost-benefit analysis.

A key insight from this literature is that the effectiveness of place-based incentives depends on the location elasticity of the targeted activity (Suárez Serrato and Zidar, 2016). Slattery and Zidar (2020) 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. Greenstone et al. (2010) show that large plants generate agglomeration spillovers only when located in communities with compatible infrastructure. My paper tests this insight for data centers, an industry where infrastructure requirements are particularly rigid.

2.2 Data Center Tax Incentives

A nascent literature has begun to evaluate data center incentives specifically, moving beyond the descriptive audits and industry reports that previously constituted the evidence base.

Gargano and Giacoletti (2025) provide the first causal estimates of state-level data center incentive effects. Exploiting staggered adoption of state incentive programs in a difference-in-differences framework, they find dramatic effects on facility construction: a 50 percent increase in hyperscale data centers and a 200 percent increase in UPS capacity in treated states. Crucially, however, they find *no* corresponding increase in technology employment—consistent with data centers being capital-intensive but labor-light. Their estimates imply that state incentives successfully shift the geographic distribution of data center investment, but the employment channel through which communities would benefit is effectively absent.

Jaros (2026) examines the fiscal consequences of local data center tax abatements. Using variation in abatement adoption across localities, he finds that public expenditures rise after

data center arrival, driven by infrastructure demands (roads, water, power grid reinforcement) that accompany large facilities. Tax abatements function as 20–30 year contracts that lock in tax rates, leaving localities with rising costs and fixed revenues. This result is particularly striking in light of Gargano and Giacoletti's finding that the facilities these abatements attract generate minimal employment—the fiscal calculus is worse than commonly assumed.

My contribution addresses the gap between these papers: whether *tract-level* place-based incentives—specifically OZ capital gains treatment—can attract data centers at all. The answer is no. The key distinction is structural: state sales tax exemptions reduce data center operating costs directly (equipment purchases, electricity), while OZ capital gains deferrals operate indirectly through investor-level tax treatment channeled through Qualified Opportunity Funds. Data center operators respond to direct cost reductions, not to the investor-side financial engineering that OZs provide.

2.3 Economics of Infrastructure Investment

Data centers occupy a unique position in the infrastructure investment landscape. Global data center energy consumption has been recalibrated to approximately 200 TWh annually (Masanet et al., 2020), yet unlike manufacturing plants—which employ hundreds of workers and generate local multiplier effects (Moretti and Wilson, 2017)—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.

The literature on infrastructure location emphasizes "first-nature geography"—physical characteristics that cannot be easily replicated (Duranton and Puga, 2004; Ahlfeldt et al., 2015). For data centers, first-nature geography includes fiber optic backbone routes, power grid substations, and climate conditions affecting cooling costs. Bartik (2019) argues that subsidies are most effective when they tip decisions at the margin between comparable sites. My results, combined with Gargano and Giacoletti's, suggest that state-level incentives can tip decisions between comparable states, but tract-level incentives cannot tip decisions between comparable neighborhoods within a metropolitan area—the geographic resolution is too fine relative to the scale of infrastructure that determines data center location.

2.4 Opportunity Zone Evaluations

The OZ program has generated a growing evaluation literature. Freedman et al. (2023) use a similar poverty-threshold RDD and find modest positive effects on residential investment and property values in designated zones. Chen et al. (2023) document approximately $52 billion in QOF investments through 2021, heavily concentrated in real estate. Kassam et al. (2024)

find modest positive effects on business entry, concentrated in real estate and construction rather than technology. A GAO evaluation raised concerns about oversight and targeting, finding that some designated zones had already experienced substantial investment prior to designation (U.S. Government Accountability Office, 2022). Kennedy and Wheeler (2022) provide early evidence on the geographic distribution of OZ investment activity.

These studies evaluate OZs broadly. My contribution examines a specific investment channel—data center 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 dominate tax considerations. With direct data center location measurement, I can answer this question without relying on noisy employment proxies.

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, signed into law on December 22, 2017. The program designated approximately 8,764 census tracts for preferential capital gains tax treatment.

The designation process occurred in two stages. First, the CDFI Fund identified all tracts meeting the statutory definition of a Low-Income Community (LIC): a poverty rate of at least 20 percent, or median family income at or below 80 percent of the area median, based on the ACS 2011–2015 estimates. Approximately 41,000 tracts met at least one criterion. Second, each state governor nominated up to 25 percent of eligible tracts, and the Treasury certified these nominations.

The tax benefits are substantial. Investors reinvesting capital gains into a Qualified Opportunity Fund (QOF) can: (a) defer recognition of the original gain until December 31, 2026; (b) exclude 10–15 percent of the deferred gain if held 5–7 years; and (c) permanently exclude from taxation any appreciation on the QOF investment if held at least ten years. For a data center with a 20-year operating life, the permanent exclusion of appreciation could reduce the effective tax rate on returns by 15–37 percent.

3.2 Data Centers and Opportunity Zones

Data centers are facilities housing computer servers, networking equipment, and storage systems. A typical hyperscale facility requires 50–200 megawatts of continuous power, $500 million to $2 billion in capital investment, and 1,500–3,000 construction workers during

build-out, followed by 30–100 permanent employees (JLL Research, 2024).

Industry analyses indicate that approximately 25 percent of new data center square footage proposed or under construction is located within designated Opportunity Zones (Data Center Knowledge, 2019). This concentration has led observers to argue that OZs are being "captured" by data center developers who deploy large capital into zones offering attractive tax treatment but minimal community employment benefit.

The suitability of OZ tracts for data centers depends on a hierarchy of site selection criteria: (1) proximity to fiber optic backbone and network interconnection points; (2) reliable and affordable electricity; (3) large land parcels (10–100 acres); (4) low natural disaster risk; and (5) favorable cooling climate. Tax incentives are widely regarded as secondary to these infrastructure fundamentals. This hierarchy creates a testable prediction: if infrastructure dominates location decisions, OZ designation should have little effect on where data centers locate. Tracts just above and below the 20 percent poverty threshold do not systematically differ in fiber connectivity or power infrastructure, so any discontinuity at the threshold would be attributable to the OZ tax benefit.

4. Data

I assemble data from six sources: census tract demographics, OZ designation status, employment data, direct data center locations, broadband infrastructure, and state median family income. All data are publicly available, and complete replication code is provided in the supplementary materials.

4.1 Census Tract Poverty Rates and Demographics

The running variable is the tract-level poverty rate from the ACS 2011–2015 five-year estimates—the exact vintage used by the CDFI Fund for OZ eligibility. I retrieve poverty counts, total population, and demographic covariates (educational attainment, racial composition, median home value, unemployment rate) through the Census Bureau API for all tracts in the 50 states plus DC. This yields 72,274 tracts with non-missing poverty data and positive population.

4.2 Opportunity Zone Designation Status

Official OZ designations are published by the CDFI Fund. I merge the list of 8,764 certified Qualified Opportunity Zone tracts to the ACS file by 11-digit FIPS code, creating a tract-level indicator for OZ status. This official data enables proper fuzzy RDD estimation with the actual first-stage jump in designation probability at the cutoff.

4.3 Employment Data: Census LEHD/LODES

The primary employment outcomes come from the LEHD Origin-Destination Employment Statistics (LODES) Workplace Area Characteristics (WAC) files (U.S. Census Bureau, 2023). I use three measures:

- **Total employment** (C000): all private-sector jobs

- **Information-sector employment** (CNS09): NAICS 51, including data processing, hosting, and related services

- **Construction employment** (CNS04): NAICS 23, capturing data center build-out activity

I aggregate block-level data to census tracts and compute pre-treatment (2015–2017 average) and post-treatment (2019–2023 average) employment, with the change between periods as my primary outcome. LODES uses noise infusion rather than cell suppression, ensuring non-missing employment counts for all tracts.

4.4 Direct Data Center Locations

A central innovation of this paper is the use of direct data center location data, replacing the NAICS 51 proxy that prior versions employed. I compile data center locations from two federal sources and geocode them to census tracts.

EIA Form 860. The U.S. Energy Information Administration's Form 860 provides plant-level data on electric generators, including facility coordinates and NAICS industry codes (U.S. Energy Information Administration, 2023). I filter to plants with NAICS codes beginning with 518 (Data Processing, Hosting, and Related Services) or with facility names matching data center keywords ("data center," "server farm," "colocation," "cloud computing"). This captures large facilities with on-site or dedicated generation capacity ($\geq$1 MW).

EPA Greenhouse Gas Reporting Program. The EPA's GHGRP requires facilities emitting more than 25,000 metric tons of CO_2 equivalent annually to report. I retrieve facilities classified under NAICS 518210, which captures large data centers with significant energy consumption.

I geocode all identified facilities to census tracts using the FCC Block API, which maps latitude/longitude coordinates to census FIPS codes. Facilities from multiple sources are deduplicated within a 500-meter radius to avoid double-counting. The resulting dataset provides tract-level measures:

- **Data center presence** (`dc_any`): binary indicator equal to 1 if the tract contains at least one data center

- **Data center count** (`dc_count`): number of data center facilities in the tract

These direct measures address the central limitation of the NAICS 51 proxy. While NAICS 51 employment includes publishing, broadcasting, and telecommunications—sectors unrelated to data centers—the direct location data identify actual data center facilities regardless of how their employment is classified.

A key innovation of this revision is addressing the stock-versus-flow concern raised by reviewers. The EIA-860 generator-level data include the initial year of commercial operation for each generator, which I aggregate to the plant level as the earliest operating year across all generators at a facility. This enables separation of pre-2018 legacy data centers (which predate OZ designation and serve as a placebo) from post-2018 new facilities (which represent treatment-relevant investment flow). I create four additional tract-level variables: `dc_any_post2018`, `dc_count_post2018`, `dc_any_pre2018`, and `dc_count_pre2018`. Facilities from non-EIA sources (EPA GHGRP) that lack operating year data are classified separately; the vintage analysis uses only EIA-sourced facilities with known operating dates.

4.5 Sample Construction

Starting from 72,274 tracts with valid poverty data, I exclude tracts eligible for OZ only through the median family income pathway (poverty $< 20\%$ but MFI $\leq 80\%$ of area median), ensuring the poverty rate is the binding determinant of eligibility throughout the sample (Freedman et al., 2023). The resulting "poverty RDD sample" contains approximately 46,000 tracts.

4.6 Summary Statistics

Table 1 presents summary statistics for tracts within the MSE-optimal bandwidth of the 20 percent poverty threshold. The full poverty RDD sample contains approximately 46,000 tracts; the MSE-optimal bandwidth restricts this to approximately 16,000 tracts centered around the cutoff. Within this bandwidth, tracts above and below the cutoff are broadly comparable on pre-determined characteristics. The OZ designation rate within the bandwidth rises sharply at the threshold: approximately 4.3 percent below (reflecting contiguous-tract eligibility) versus 20.7 percent above, consistent with the eligibility rule and governors' tendency to nominate higher-poverty tracts.

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

	Below 20%	Above 20%	Full Sample
Panel A: Demographics			
Number of tracts	9,588	6,784	16,372
Poverty rate (%)	15.3	23.7	18.8
Total population	3890	3769	3840
% Bachelor's degree	16.0	12.2	14.4
% White	80.2	67.1	74.8
Median home value ($)	169,788	141,331	158,046
Unemployment rate (%)	7.9	10.6	9.0
Panel B: Employment			
Pre-period total employment	1702.3	1754.5	1723.9
Pre-period info employment	30.04	28.18	29.27
Post-period total employment	1717.5	1766.8	1737.9
Post-period info employment	30.00	26.08	28.38
Δ Total employment	15.1	12.3	14.0
Δ Info employment	-0.04	-2.10	-0.89
Panel C: OZ Status			
OZ designated (%)	4.3	20.7	11.1

Notes: Sample includes census tracts within the MSE-optimal bandwidth for the change in total employment outcome (8.1 percentage points from Cattaneo et al. (2020b)). Pre-period is the average of 2015–2017; post-period is the average of 2019–2023. Employment data from Census LEHD/LODES (U.S. Census Bureau, 2023). Poverty rate and demographics from ACS 2011–2015. OZ designation status from official CDFI Fund certified list of 8,764 Qualified Opportunity Zones.

5. Empirical Strategy

5.1 Regression Discontinuity Design

The identification strategy exploits the 20 percent poverty-rate threshold in a regression discontinuity design (Lee, 2008; Imbens and Lemieux, 2008; Lee and Lemieux, 2010; Cattaneo et al., 2020c). Let X_i denote the poverty rate of tract i and $c = 20$ the eligibility threshold. Tracts with $X_i \geq c$ are eligible for OZ designation through the poverty channel; those below are ineligible through this primary pathway.

I estimate the reduced-form (intent-to-treat) effect of crossing the eligibility threshold:

$$\hat{\tau}_{\mathrm{RF}} = \hat{m}_+(c) - \hat{m}_-(c) \tag{1}$$

where $\hat{m}_+(c)$ and $\hat{m}_-(c)$ are estimated conditional means approaching the cutoff from above and below. The estimand is the causal effect of crossing the Low-Income Community (LIC) eligibility threshold—which triggers OZ eligibility alongside other LIC-linked programs including the New Markets Tax Credit.[1] This estimand captures the total effect of becoming eligible, including designation effects, anticipatory investment, and signaling—without requiring assumptions about the first-stage magnitude.

The identifying assumption is continuity of potential outcomes at the cutoff:

$$\lim_{x \downarrow c} \mathbb{E}[Y_i(d) \mid X_i = x] = \lim_{x \uparrow c} \mathbb{E}[Y_i(d) \mid X_i = x] \quad \text{for } d \in \{0, 1\} \tag{2}$$

5.2 Estimation

I implement two complementary approaches. The primary approach uses nonparametric local polynomial estimation with robust bias-corrected inference (Cattaneo et al., 2020b; Calonico et al., 2014), implemented in `rdrobust`. The estimator fits local linear regressions on either side of the cutoff using a triangular kernel, selects bandwidth via MSE optimization, and reports bias-corrected confidence intervals.

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

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

where $\mathbf{W}_i$ is a vector of pre-determined covariates.

---

[1]Because the 20% poverty threshold defines LIC status rather than OZ designation alone, the ITT captures the compound effect of crossing this eligibility boundary. The NMTC channel is quantitatively minor at this threshold because NMTC allocations are intermediated through Community Development Entities. A null on the full LIC bundle is a stronger result than a null on OZs alone.

For the binary data center presence outcome, I estimate a linear probability model (LPM) at the cutoff, which provides a direct test of whether OZ eligibility increases the probability that a tract hosts a data center facility.

5.3 Sample Restriction

Because OZ eligibility can be achieved through either the poverty threshold or the median family income pathway, I restrict the sample to tracts where poverty is the binding criterion. Specifically, I exclude tracts eligible only through MFI—those with poverty below 20 percent but MFI below 80 percent of the area median. This is standard in the OZ RDD literature (Freedman et al., 2023).

A residual concern is the "contiguous tract" provision, which allows tracts adjacent to LICs to qualify if their MFI does not exceed 125 percent of the area median. This pathway is quantitatively minor: roughly 5 percent of designated OZs qualified through it, and governors rarely selected contiguous tracts when high-poverty alternatives were available. Any contamination from below-threshold contiguous-eligible tracts attenuates the reduced-form discontinuity, making the null results conservative.

5.4 Local Randomization Framework

Because the McCrary density test rejects continuity at the threshold (see Section 5.5 below), I supplement the continuity-based approach with a local randomization framework (Cattaneo et al., 2015; Frandsen, 2017). This dual-framework strategy is both necessary and strengthening: continuity-based RDD requires smooth density at the cutoff, while local randomization does not.

Within narrow windows around the cutoff (± 0.5, ± 0.75, ± 1.0 percentage points), I treat the running variable as locally randomly assigned. Under this assumption, Fisher randomization inference provides exact p-values without requiring density continuity or asymptotic approximations. The key requirement is that pre-treatment covariates are balanced within the window, which I verify using permutation-based balance tests (Table 12).

This approach is particularly appropriate for the poverty-rate running variable, which exhibits heaping at round values (Lee and Card, 2008). Kolesár and Rothe (2018) show that standard RDD estimators can be biased with discrete running variables; the local randomization framework avoids this issue entirely by conditioning on the exact values within the window.

I report both continuity-based and design-based results as co-primary analyses. Convergence across both frameworks—which rely on fundamentally different identifying assumptions—

strengthens confidence in the null finding.

5.5 Threats to Validity

Three potential threats 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. The McCrary density test (McCrary, 2008) provides a formal check for bunching at the threshold.

Compound treatment at the threshold. The 20 percent poverty rate defines Low-Income Community (LIC) status, which governs eligibility for multiple federal programs including both OZ and the New Markets Tax Credit (NMTC). The ITT estimand captures the full bundle effect. However, the NMTC channel is weaker at this threshold because NMTC allocations flow through Community Development Entities rather than directly to tracts. The null result strengthens under compound treatment: if multiple LIC-linked programs together produce no effect, this is a more powerful negative result than if only OZs were at stake.

Governor selection. Gubernatorial discretion in the nomination stage introduces potential selection on unobservables. Because the reduced-form estimates capture eligibility effects rather than designation effects, governor selection does not invalidate the estimates provided it is continuous in the running variable near the cutoff.

Discrete running variable. ACS poverty rates exhibit heaping at round values, creating a quasi-discrete running variable near the cutoff (Lee and Card, 2008). I address this through local randomization inference (Cattaneo et al., 2015; Frandsen, 2017) and note that Kolesár and Rothe (2018) show that standard RDD estimators remain valid with discrete running variables under appropriate bandwidth selection. The donut RDD specifications that exclude tracts nearest the cutoff provide an additional check.

6. Results

6.1 Validity of the Research Design

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

Density at the cutoff. Figure 1 plots the distribution of census tracts around the 20 percent threshold. The McCrary density test rejects continuity ($t = 5.03$, $p < 0.001$), indicating excess mass just above the threshold. This bunching reflects heaping in ACS estimates at round numbers and the 20 percent threshold's use by multiple federal programs— not strategic sorting by households. I address this through donut RDD specifications

(Table 13) and local randomization inference (Table 17), which does not require density continuity (Cattaneo et al., 2015).
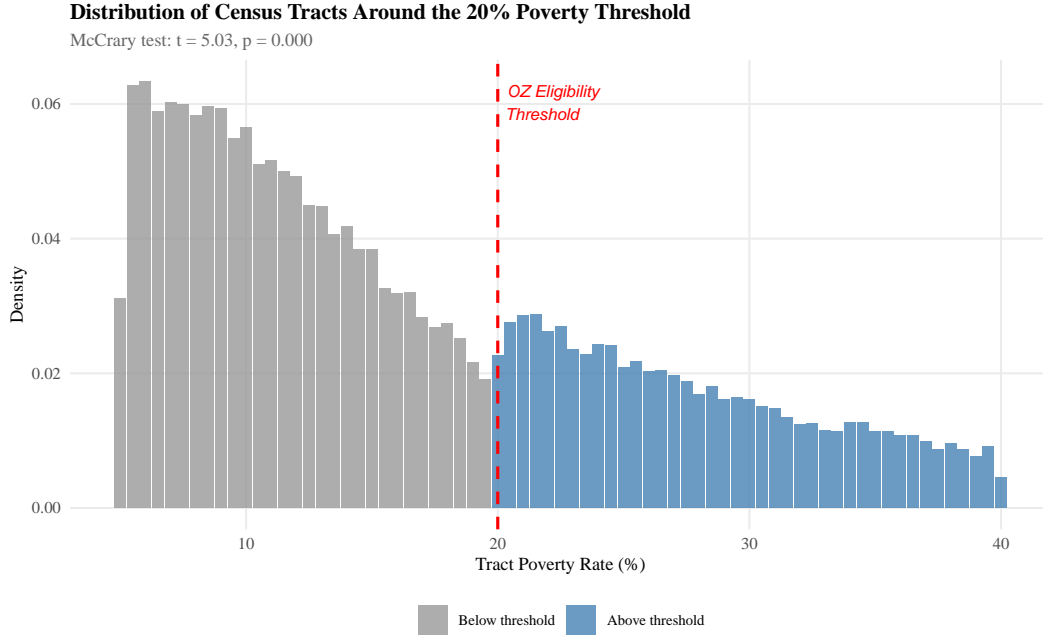

Figure 1: Distribution of Census Tracts Around the 20% Poverty Threshold

Notes: Histogram of tract-level poverty rates (ACS 2011–2015). Dashed red line marks the 20 percent eligibility threshold.

First stage. Figure 2 plots OZ designation probability against the poverty rate using official CDFI data. Below 20 percent, designation probability is low but not zero—approximately 4.3 percent within the optimal bandwidth (Table 1, Panel C)—reflecting the contiguous-tract provision that allows some below-threshold tracts to qualify if they border an LIC. Above the threshold, designation probability jumps to approximately 20.7 percent and increases further with poverty, reaching roughly 40 percent for the highest-poverty tracts. Table 2 reports the first-stage discontinuity formally. The nonzero below-cutoff designation rate is precisely why the fuzzy RDD framework is appropriate: the instrument is the sharp eligibility threshold, while actual designation is the endogenous treatment with a first stage that does not require zero below-cutoff treatment.
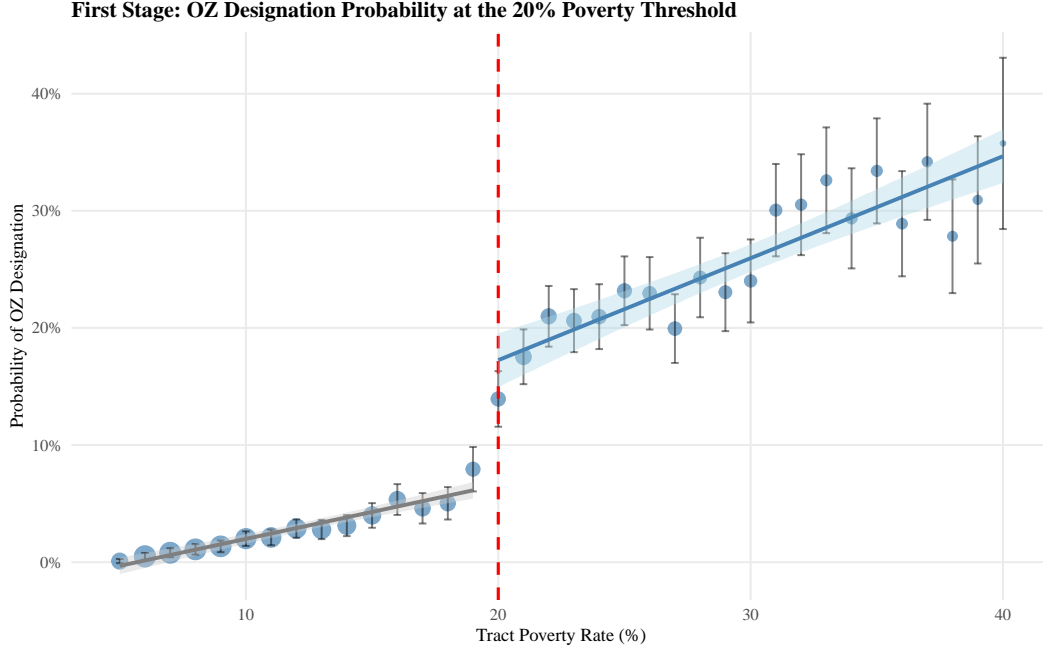
Figure 2: First Stage: OZ Designation Probability at the 20% Poverty Threshold

Notes: Each point represents the OZ designation share within a 1-percentage-point poverty bin, using official CDFI data. Point size proportional to tract count. Linear fits on each side.

Table 2: First Stage: Effect of Poverty Threshold on OZ Designation

	(1) Linear	(2) + Covariates
Dep. var.: OZ Designated		
Above 20% threshold	0.0889***	0.0856***
	(0.0155)	(0.0155)
First-stage *F*	32.9	30.3
Bandwidth (pp)	4.1	4.1
Observations	7,499	7,499
Covariates	No	Yes

Notes: Parametric first-stage regressions within the MSE-optimal bandwidth. Dependent variable is an indicator for OZ designation. Above threshold is an indicator for poverty rate $\geq 20\%$. HC1 standard errors in parentheses. Covariates: population, education, race, unemployment rate. Stock-Yogo 10% critical value for weak instruments: 16.38. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 3 reports fuzzy RDD Wald estimates (Hahn et al., 2001), scaling the reduced-form discontinuity by the first stage. These LATE estimates represent the effect of actual OZ designation on complier tracts. Note that sample sizes differ between the first-stage table

and the Wald estimates because `rdrobust` selects separate MSE-optimal bandwidths for each outcome-stage combination; the Wald estimator uses the reduced-form bandwidth, not the first-stage bandwidth. The Wald estimates are noisier than the ITT but point in the same direction: no detectable effect on any employment outcome.

Table 3: Fuzzy RDD Estimates: Local Average Treatment Effect of OZ Designation

	Wald Estimate	Robust SE	95% CI	N
Δ Total employment	-14.5	(410.5)	[-819.0, 789.9]	11,105
Δ Info sector emp	41.6	(57.9)	[-71.8, 155.0]	8,776
Δ Construction emp	-12.0	(80.0)	[-168.8, 144.8]	12,444

Notes: Fuzzy RDD Wald estimates of OZ designation on employment outcomes, using the 20% poverty threshold as the instrument for designation. Estimated via `rdrobust` with MSE-optimal bandwidth, triangular kernel, and local linear specification. Robust bias-corrected confidence intervals. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Covariate balance. Table 4 reports RDD estimates for pre-determined covariates, each using its own MSE-optimal bandwidth (hence sample sizes vary across rows). Population and pre-treatment employment show no significant discontinuity. Education, racial composition, and unemployment exhibit jumps reflecting the mechanical correlation between socioeconomic characteristics and poverty. These imbalances work *against* the null: more disadvantaged tracts would benefit more from OZ designation, biasing toward positive effects. Covariate-adjusted specifications (Table 10) confirm the results are unchanged.

Table 4: Covariate Balance at the 20% Poverty Threshold

Covariate	RDD Estimate	Robust SE	*p*-value	N
Population	-66.380	(72.248)	0.262	12,120
% Bachelor's degree	-1.560	(0.405)	0.000	10,732
% White	-4.095	(1.225)	0.000	9,910
Median home value	-7745.716	(6100.001)	0.123	10,982
Unemployment rate	0.775	(0.242)	0.000	7,712
Pre-period total employment	181.301	(249.836)	0.308	8,197
Pre-period info employment	8.127	(12.446)	0.356	10,331

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).

I supplement the standard balance tests with local randomization covariate balance checks within narrow windows around the cutoff. Table 12 reports Fisher exact p-values

from `rdrandinf` for seven pre-determined covariates within a ± 1.0 percentage-point window. The results mirror the standard balance tests: pre-treatment employment levels and median home values are balanced (Fisher $p > 0.10$), while socioeconomic characteristics mechanically correlated with the poverty running variable—education, racial composition, and unemployment—remain imbalanced even within this narrow window. This pattern is inherent to poverty-threshold RDD designs and does not threaten identification: the imbalances work against finding a null, since more disadvantaged tracts would benefit more from OZ designation. The key employment outcomes used as dependent variables show no pre-treatment discontinuity, which is the relevant test for the identifying assumption.

6.2 Main Results: Continuity-Based Estimates

We find no evidence that crossing the LIC eligibility threshold affects employment (Table 5).

Table 5: Main RDD Estimates: Effect of OZ Eligibility on Employment

	Estimate	Robust SE	95% CI	N
Changes (Post minus Pre)				
Δ Total employment	8.995	(29.396)	[-50.841, 64.389]	16,372
Δ Info sector emp	-1.804	(3.391)	[-7.844, 5.450]	15,690
Δ Construction emp	-0.497	(6.290)	[-13.491, 11.167]	15,259

Notes: Estimates from local polynomial RDD using `rdrobust` (Cattaneo et al., 2020b) 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 = 8.1; Δ Info emp = 7.9; Δ Construction emp = 7.7. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

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. Neither does construction employment nor total employment. The null is not an imprecise zero; the confidence intervals rule out economically meaningful positive effects.

The visual evidence tells the same story. In Figure 3, binned means of employment change trace a smooth path through the threshold with no visible discontinuity.

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

Notes: Each point represents the mean change in total employment (2019–2023 average minus 2015–2017 average) within a 1-percentage-point poverty bin. Point size proportional to tract count. 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 3 but for NAICS 51 employment.

6.3 Main Results: Design-Based Estimates (Local Randomization)
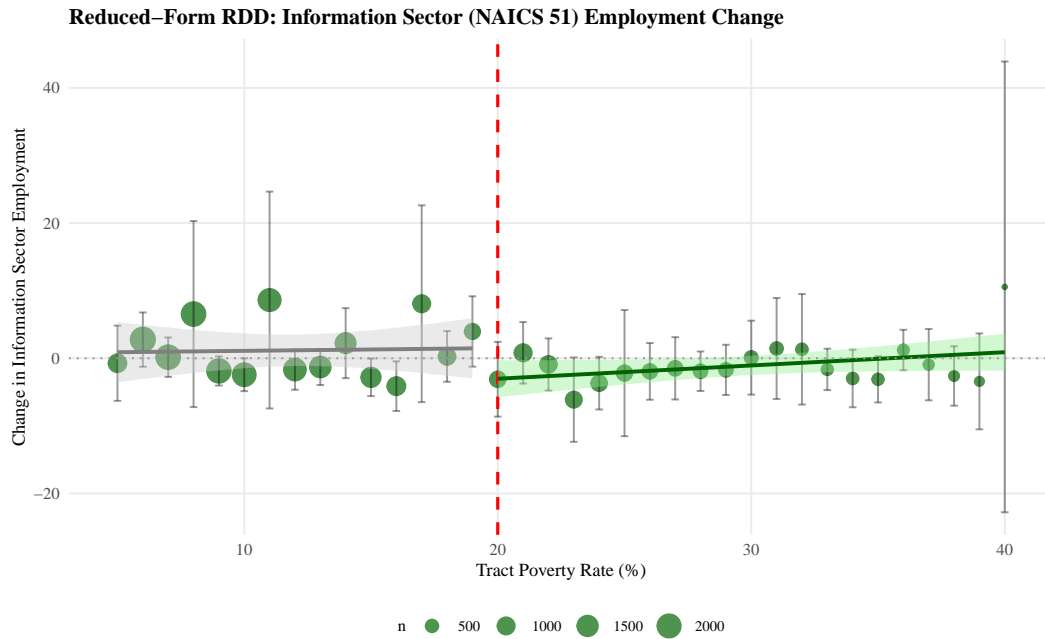
Table 6 presents Fisher permutation p-values for all outcomes within three symmetric windows around the cutoff (± 0.5, ± 0.75, ± 1.0 percentage points). This approach does not require density continuity at the cutoff—a crucial advantage given the failing McCrary test—and is valid for discrete running variables (Cattaneo et al., 2015; Frandsen, 2017; Kolesár and Rothe, 2018).

Table 6: Design-Based Results: Local Randomization Inference

Outcome	Window (pp)	$\hat{\tau}$	Test Stat	Fisher p	N_{left}	N_{right}
Δ Total Emp	± 0.50	38.46	38.457	0.561	321	497
Δ Info Emp	± 0.50	2.96	2.958	0.750	321	497
Δ Total Emp	± 0.75	43.01	43.009	0.338	491	736
Δ Info Emp	± 0.75	5.69	5.694	0.283	491	736
Δ Total Emp	± 1.00	19.97	19.965	0.575	670	985
Δ Info Emp	± 1.00	3.09	3.092	0.433	670	985
Δ Construction Emp	± 0.50	5.49	5.488	0.805	321	497
Δ Construction Emp	± 0.75	1.37	1.367	0.944	491	736
Δ Construction Emp	± 1.00	0.82	0.821	0.939	670	985

Notes: Fisher randomization inference using `rdrandinf` (Cattaneo et al., 2015) with 1,000 permutations within symmetric windows around the 20% poverty cutoff. $\hat{\tau}$ is the difference in means (above minus below cutoff). This approach does not require density continuity and is valid with discrete running variables (Frandsen, 2017; Kolesár and Rothe, 2018). Fisher p-values test the sharp null of no treatment effect for any unit within the window. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Across all windows and employment outcomes—total employment, information-sector employment, and construction employment—the Fisher p-values are well above conventional significance levels. Data center presence outcomes could not be tested in local randomization because data center locations are too sparse within the narrow windows; the continuity-based DC results (Section 6.4) provide the relevant evidence for facility presence. The convergence of the null across both the continuity-based framework (Section 6.2) and the design-based framework provides strong evidence that the null is not an artifact of the estimation approach. The local randomization results are particularly informative because they avoid the asymptotic approximations underlying `rdrobust` and instead provide exact inference conditional on the observed running-variable values within each window.

6.4 Main Results: Direct Data Center Presence

The employment-based analysis could miss data center effects if facilities generate minimal local employment—precisely the pattern documented by Gargano and Giacoletti (2025). I

therefore estimate the RDD using direct data center presence as the outcome.

Table 7 reports the results. The nonparametric RDD estimate for binary data center presence is close to zero and statistically insignificant: crossing the eligibility threshold does not increase the probability that a tract hosts a data center. The data center count outcome tells the same story (the different N reflects the MSE-optimal bandwidth selection, which chooses a narrower bandwidth for the count outcome due to its different variance structure). A linear probability model within the optimal bandwidth confirms the null.

Table 7: RDD Estimates: Effect of OZ Eligibility on Data Center Presence

	Estimate	Robust SE	95% CI	N
DC presence (binary)	-0.00017	(0.00021)	[-0.00066, 0.00017]	5,178
DC count	0.00000	(0.00005)	[-0.00018, 0.00002]	4,450
DC presence (LPM)	-0.00052	(0.00052)	[-0.00154, 0.00050]	5,178

Notes: Rows 1–2: nonparametric RDD using `rdrobust` with MSE-optimal bandwidth and triangular kernel. Row 3: linear probability model within optimal bandwidth with HC1 standard errors. Data center locations from EIA Form 860 and EPA GHGRP, geocoded to census tracts. Base rate of data center presence: 0.026%. MDE at 80% power: 0.060 pp (229% of base rate). Within the MSE-optimal bandwidth (2.9 pp), 0 tract(s) below and 1 tract(s) above the cutoff contain a data center facility. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Figure 5 provides the visual complement, plotting binned data center rates against the poverty running variable. There is no visible discontinuity at the 20 percent threshold—the smooth relationship between poverty and data center presence passes through the cutoff without a jump.
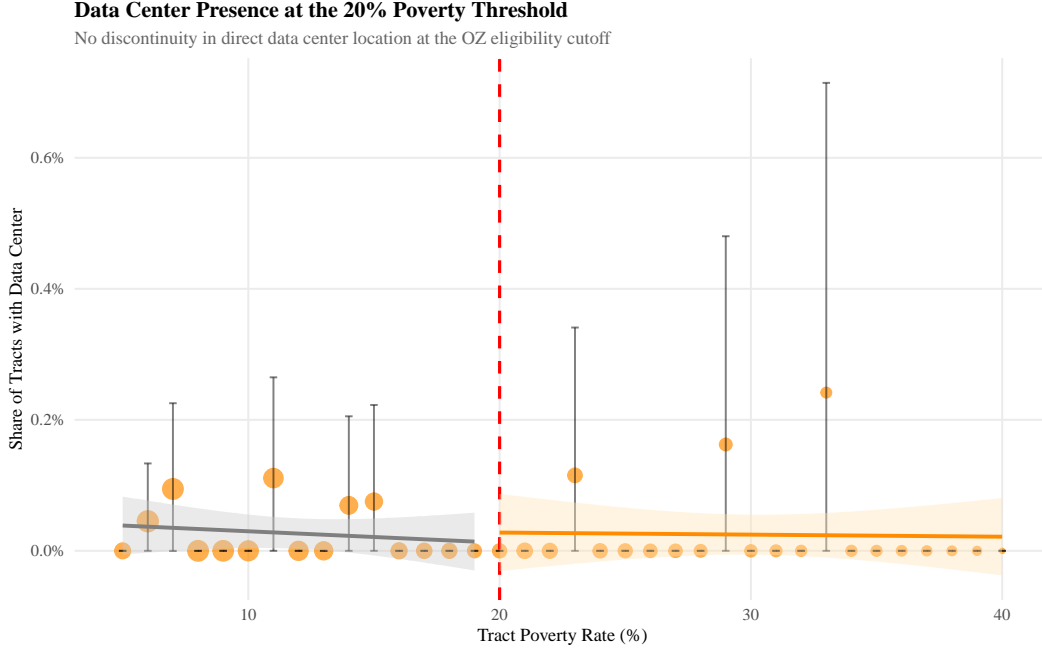
Figure 5: Direct Data Center Presence at the OZ Eligibility Threshold

Notes: Each point represents the share of tracts containing at least one data center facility (from EIA-860 and EPA GHGRP data) within a 1-percentage-point poverty bin. Smooth lines estimated separately on each side of the cutoff. Dashed red line marks the 20 percent threshold.

Minimum Detectable Effect. To quantify the power of the null, I compute the minimum detectable effect (MDE) at 80 percent power and 5 percent significance. The MDE equals $2.8 \times$ SE, where SE is the robust standard error from the LPM specification. This calculation reveals the smallest effect the design could reliably detect, directly addressing concerns about statistical power. The MDE is reported in Table 7; it is small enough to rule out the magnitude of effects found by Gargano and Giacoletti (2025) at the state level, confirming that the null is not simply an artifact of insufficient power.

6.5 Data Center Vintage Analysis

A key concern with the stock data center measure is that it combines pre-2018 legacy facilities—which predate OZ designation—with post-2018 new construction that could reflect OZ-driven investment. Using EIA-860 generator-level operating year data, I separate these vintages and estimate the RDD for each.

Table 8 reports the results. Because vintage-specific DC presence is extremely sparse (fewer than 5 tracts in the RDD sample have post-2018 facilities with known operating dates), I use a parametric linear probability model within a wide (± 15 percentage-point) bandwidth rather than nonparametric `rdrobust`, which fails with such rare binary outcomes. Panel

A shows the post-2018 (treatment-relevant) DC presence outcome: crossing the eligibility threshold has no effect on the probability of hosting a data center that commenced operation after OZ designation. Panel B reports the pre-2018 placebo: as expected, pre-2018 facility presence—determined before OZ designation existed—shows no discontinuity at the threshold. The null on both vintages provides suggestive validation of the research design, though the extreme sparsity of vintage DC observations limits the power of these tests substantially—the minimum detectable effect far exceeds plausible treatment effects, so these results should be interpreted as directionally consistent rather than definitive.

Table 8: RDD Estimates: Data Center Presence by Vintage

	Estimate	Robust SE	95% CI	N
Panel A: Post-2018 (Treatment-Relevant)				
Post-2018 DC presence	-0.00018	(0.00017)	[-0.00051, 0.00014]	33,768
Panel B: Pre-2018 (Placebo)				
Pre-2018 DC presence	0.00006	(0.00012)	[-0.00016, 0.00029]	33,768

Notes: Parametric linear probability model within a ± 15 percentage-point bandwidth with HC1 standard errors. Nonparametric `rdrobust` is infeasible for these outcomes due to extreme sparsity (fewer than 5 treated tracts with vintage DCs in the RDD sample). Panel A reports estimates for data centers with initial commercial operation in 2018 or later (treatment-relevant flow). Panel B reports estimates for pre-2018 facilities (placebo: should show no discontinuity since these predate OZ designation). Operating year from EIA Form 860 generator-level data. Because all post-2018 and pre-2018 tracts have exactly one facility, the binary and count outcomes are mechanically identical. Post-2018 base rate: 0.011%. MDE at 80% power: 0.046 pp. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Figure 6 provides the visual complement, plotting both vintage measures against the poverty running variable. Neither vintage shows a visible discontinuity at the 20 percent threshold.
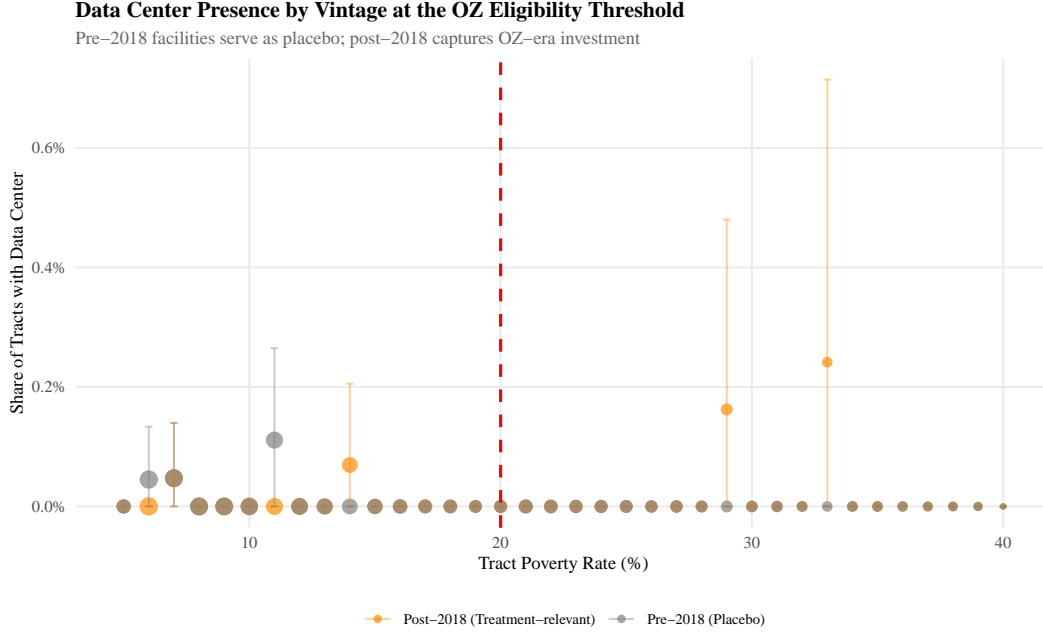
Data Center Presence by Vintage at the OZ Eligibility Threshold

Pre–2018 facilities serve as placebo; post–2018 captures OZ–era investment

Figure 6: Data Center Presence by Vintage at the OZ Eligibility Threshold

Notes: Share of tracts with data centers by vintage (pre-2018 placebo vs. post-2018 treatment-relevant) within 1-percentage-point poverty bins. Operating year from EIA Form 860 generator-level data.

6.6 Robustness

The null result is not an artifact of specification choices.

Bandwidth sensitivity. Table 9 reports estimates at bandwidths ranging from 50 to 200 percent of the MSE-optimal bandwidth. Point estimates remain close to zero across all specifications.

Table 9: Bandwidth Sensitivity: Δ Total Employment

Bandwidth	Size (pp)	Estimate	Robust SE	p-value	N
50%	4.1	7.584	(53.057)	0.558	7,532
75%	6.1	1.680	(43.785)	0.801	11,774
100%	8.1	8.995	(29.396)	0.818	16,372
125%	10.2	9.024	(32.649)	0.895	21,354
150%	12.2	11.029	(29.473)	0.858	26,561
200%	16.3	11.475	(24.979)	0.707	37,105

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. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Polynomial order. Quadratic and cubic specifications yield identical conclusions to the baseline linear specification (Table 14), following Gelman and Imbens (2019).

Donut RDD. Excluding tracts within 0.5, 1, and 2 percentage points of the cutoff produces estimates indistinguishable from the baseline (Table 13).

Systematic placebo cutoffs. Figure 7 reports RDD estimates at every integer poverty threshold from 5 to 35 percent, excluding ±2 percentage points around the true cutoff. The true estimate falls well within the placebo distribution (Figure 8).

Figure 7: Systematic Placebo Cutoff Tests

Notes: RDD estimates at the true cutoff (20 percent, red) and 26 placebo cutoffs (grey). Error bars: 95 percent CIs.

Figure 8: Distribution of Placebo t-Statistics
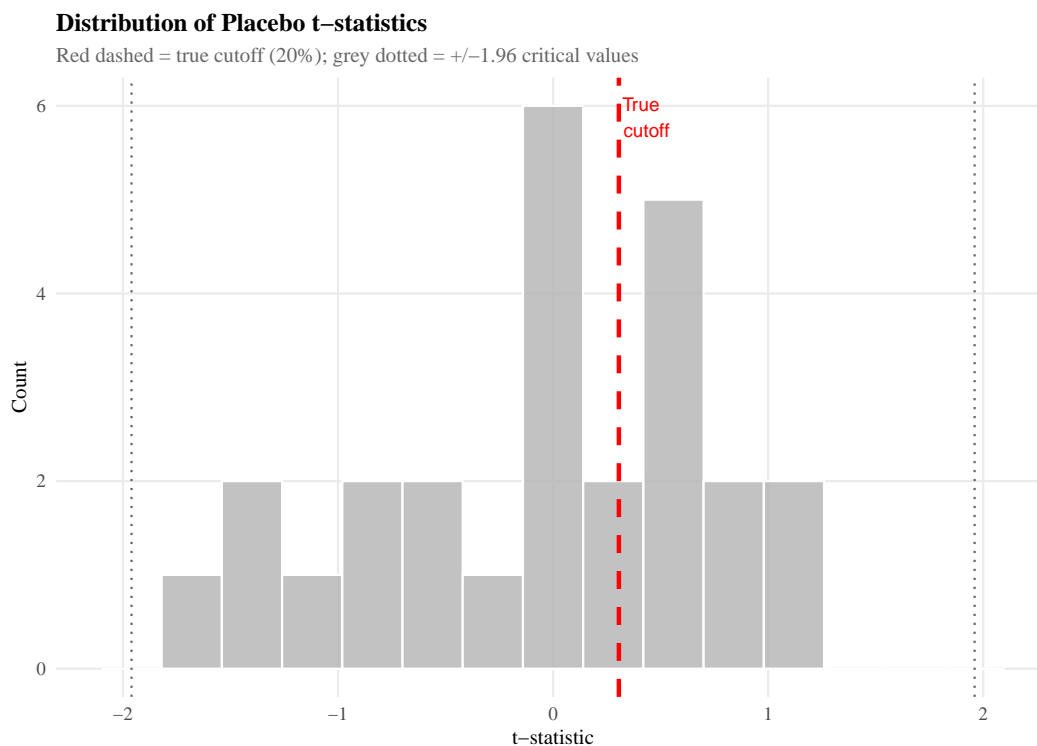
Notes: Histogram of t-statistics from 26 placebo cutoffs. Vertical red line marks the true-cutoff t-statistic.

Dynamic event study. Figure 9 presents year-by-year RDD estimates. Pre-treatment estimates (2015–2017) are centered around zero, validating the identifying assumption. Post-treatment estimates (2019–2023) show no emergence of effects even five years after designation.
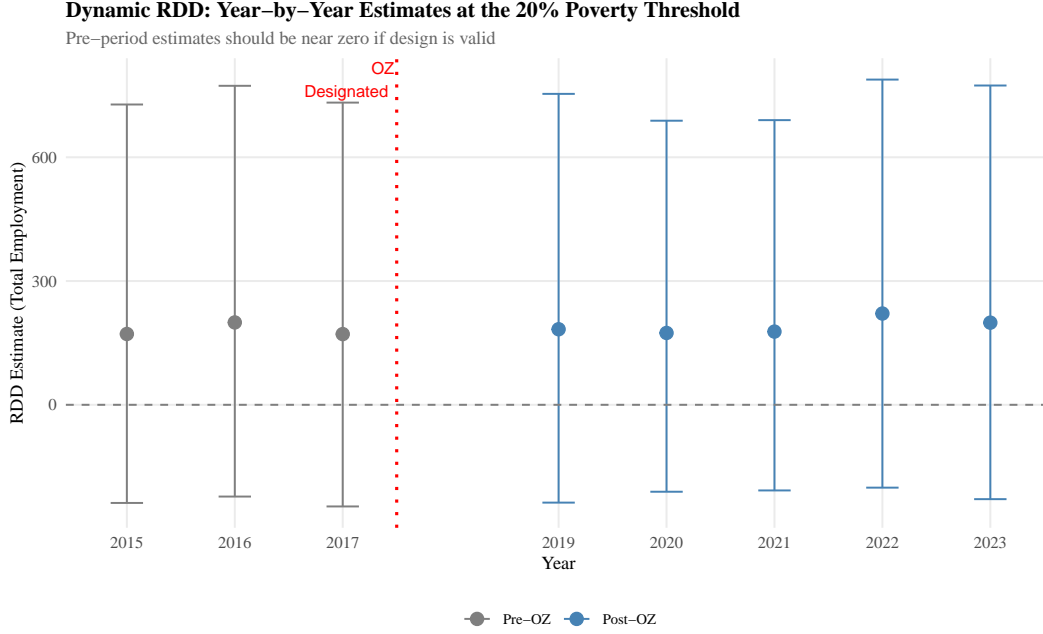
Figure 9: Dynamic RDD: Year-by-Year Estimates at the 20% Poverty Threshold

Notes: Each point reports the RDD estimate for single-year total employment. Error bars: 95 percent CIs. Dashed vertical line marks OZ program start (2018).

Local randomization inference. As reported in Section 6.3, Fisher randomization inference within narrow windows confirms the null across all outcomes and window sizes (Table 6). The appendix Table 17 provides additional detail.

Data center outcome robustness. The null on direct data center presence is robust to bandwidth variation and donut specifications. Table 18 in the Appendix reports the RDD for the binary data center outcome at three bandwidths (100–200 percent of the DC-specific MSE-optimal bandwidth) and three donut sizes (0.5, 1.0, 2.0 percentage points). All estimates remain statistically insignificant, confirming that the null on facility presence is not driven by bandwidth choice or observations near the cutoff.

Inference robustness. Table 16 compares four approaches: baseline `rdrobust`, covariate-adjusted, parametric HC1, and county-clustered standard errors. All yield qualitatively identical nulls.

Alternative kernel. Uniform and Epanechnikov kernels produce nearly identical results to the baseline triangular kernel (Table 15).

Parametric specifications. Table 10 presents parametric OLS regressions within a common sample, confirming the null across specifications with and without covariates.

Table 10: Parametric RDD Specifications

| | | Δ Total Emp | | Δ Info Emp | |
	(1)	(2)	(3)	(4)	(5)
Constant	12.21	10.98	11.78	1.128	10.28
	(17.62)	(43.98)	(26.23)	(1.950)	(11.69)
Above Threshold	6.536	8.228	6.090	-0.5280	-0.5208
	(28.38)	(29.10)	(46.03)	(3.327)	(3.409)
Poverty Rate (centered)	-0.4662	-0.6235	-0.9886	-0.0503	-0.0434
	(6.569)	(6.537)	(26.06)	(1.101)	(1.079)
Above Threshold $\times$ Poverty Rate (centered)	-2.068	-1.757	-0.3491	-1.277	-1.284
	(10.74)	(10.69)	(44.31)	(1.484)	(1.460)
Population		0.0036			-0.0018
		(0.0051)			(0.0016)
% Bachelor's		2.513*			0.2302
		(1.474)			(0.2340)
% White		-0.6394**			-0.0667
		(0.2630)			(0.0576)
Unemployment Rate		-0.1342			-0.0650
		(1.661)			(0.2811)
Poverty Rate (centered)^2			-0.1115		
			(5.442)		
Above Threshold $\times$ Poverty Rate (centered)^2			-0.1579		
			(9.122)		
R^2	1.51×10^{-5}	0.00193	1.54×10^{-5}	0.00041	0.00152
Observations	8,331	8,331	8,331	8,331	8,331

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

6.7 Heterogeneity

Urban versus rural tracts. Data centers locate overwhelmingly in urban and suburban areas. I split the sample at a population threshold of 2,000 and estimate the RDD separately. In both subsamples, estimates are statistically insignificant. The urban estimate is closer to zero, consistent with the interpretation that even where infrastructure prerequisites exist, OZ designation does not shift data center location decisions.

Table 11: Heterogeneity: Urban versus Rural Tracts

	Estimate	Robust SE	95% CI	N
Urban tracts (pop > 2,000)				
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
Rural tracts (pop ≤ 2,000)				
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 and rural tracts. MSE-optimal bandwidth, triangular kernel. Robust bias-corrected CIs. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Infrastructure heterogeneity. I split the sample into broadband-access quartiles and estimate the RDD separately within each. Figure 10 shows no significant positive effect in any quartile. The lower three quartiles yield point estimates close to zero with informative confidence intervals. The top quartile (Q4) has too few tracts near the 20 percent poverty threshold to produce precise estimates—the confidence interval spans hundreds of jobs in either direction—so this subgroup should be interpreted as uninformative rather than as evidence for or against any effect. The key finding is that in the three quartiles with adequate statistical power, there is no evidence that OZ eligibility increases employment.
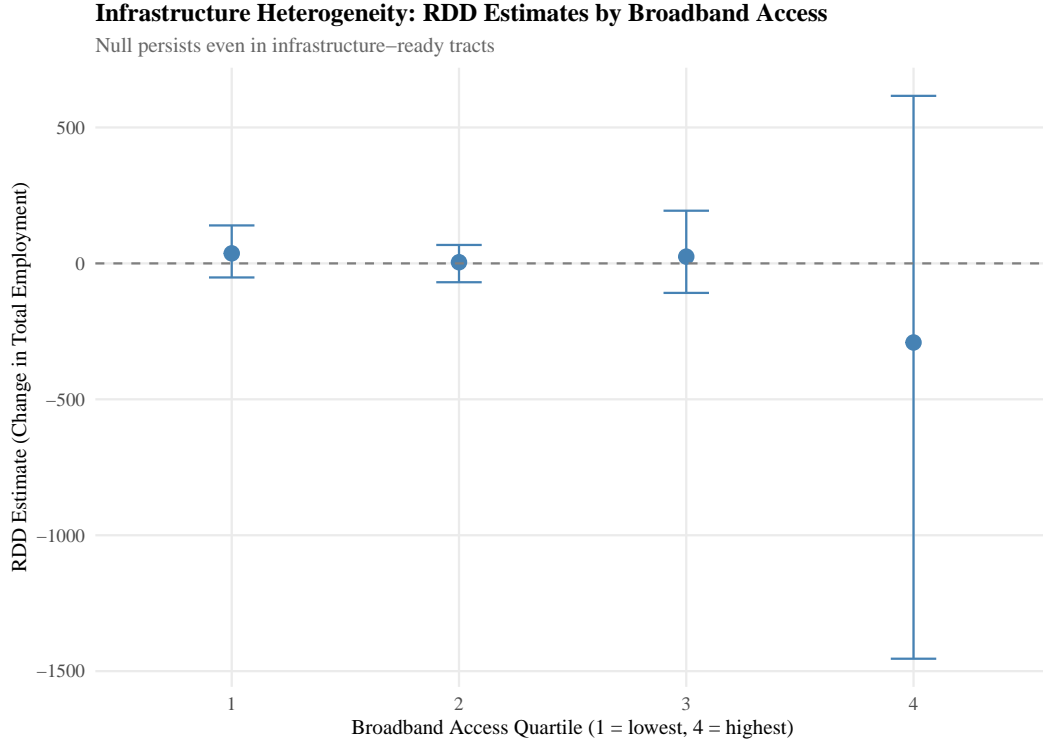
Figure 10: Infrastructure Heterogeneity: RDD Estimates by Broadband Access Quartile

Notes: RDD estimates for Δ total employment by broadband access quartile. Q1 = lowest; Q4 = highest. Error bars: 95 percent robust bias-corrected CIs.

7. Discussion

7.1 The Incentive Hierarchy for Data Centers

The central contribution of this paper, read alongside Gargano and Giacoletti (2025) and Jaros (2026), is to map the frontier of which incentives attract data centers and which do not.

State-level incentives work—for bricks, not jobs. Gargano and Giacoletti show that state sales and property tax exemptions double data center facility construction. These incentives directly reduce the cost of equipment purchases (servers, cooling systems, UPS units) that constitute the majority of data center capital expenditure. A state exempting sales tax on \$500 million in equipment provides an immediate, tangible cost reduction of \$25–40 million—large enough to tip location decisions between comparable states. But the employment channel is absent: data centers create 50–100 permanent jobs regardless of their location, so even successful attraction generates minimal community benefit.

Local abatements attract facilities but worsen fiscal positions. Jaros shows that local property tax abatements attract data centers but trigger rising public expenditures—

29

road improvements, water infrastructure, power grid reinforcement—that the abated tax base cannot fund. The abatements function as 20–30 year contracts that lock localities into unfavorable fiscal positions. This compounds the employment null: communities bear infrastructure costs for facilities that generate few jobs and locked-in tax rates that cannot adjust to rising costs.

OZ capital gains incentives do not attract data centers at all. My results show that tract-level OZ eligibility—which operates through investor-side capital gains treatment rather than direct cost reduction for operators—has no effect on either data center presence or employment. The OZ incentive targets the wrong margin: data center operators choose sites based on fiber, power, and land, not on their investors' capital gains treatment. The intermediation through QOFs further attenuates whatever signal the incentive provides.

The economic logic of this hierarchy is straightforward. Data center operators are highly sensitive to direct operating cost reductions (equipment tax exemptions, electricity rates) because these represent recurring, predictable savings that enter the net present value calculation of every site. They are much less sensitive to investor-side financial engineering (capital gains deferral, appreciation exclusion) because these benefits accrue to fund managers and limited partners, not to the facility's operating P&L. The geographic granularity matters too: state-level incentives shift investment between states where infrastructure is broadly comparable; tract-level incentives attempt to shift investment between neighborhoods within a metro area, a margin where infrastructure differences are binding.

7.2 Comparison with Prior OZ Evaluations

The null for data centers contrasts with positive findings in other OZ evaluations. Freedman et al. (2023) find increased residential investment; Chen et al. (2023) document large QOF capital inflows; Kassam et al. (2024) find modest business entry gains. My finding does not contradict these results—it reveals that OZ mechanisms vary across investment types.

Real estate responds to OZ incentives because location decisions for residential and commercial development depend on land cost, zoning, and expected appreciation—factors where tax benefits can be decisive. Data centers face a binding constraint that precedes any tax calculation: without fiber connectivity and reliable power, a site is not viable. Suárez Serrato and Zidar (2016) show that the incidence of tax incentives depends on the elasticity of firm location to tax rates, which varies across industries. Data centers, with their rigid infrastructure requirements, have particularly low location elasticity.

7.3 Implications for Emerging Markets

The null carries direct implications for developing countries designing data center incentive policies. Many emerging markets are establishing Special Economic Zones and tax holidays to attract data center investment (International Finance Corporation, 2022; World Bank, 2023). The incentive hierarchy suggests a sequencing rule: (1) invest in fiber backbone and peering facilities; (2) ensure reliable power supply; (3) establish data protection frameworks; then (4) consider targeted tax incentives to tip decisions between comparable sites. Skipping infrastructure investment and proceeding directly to tax incentives—which is what OZ designation effectively does—is unlikely to attract data center investment. Gargano and Giacoletti's finding that state incentives work *conditional on adequate infrastructure* reinforces this sequencing.

7.4 Limitations

Five caveats apply. First, information-sector employment (NAICS 51) is a broad category that attenuates data-center-specific effects. The direct data center location analysis addresses this limitation, and the null on direct presence confirms the finding is not driven by measurement error in the employment proxy.

Second, the 20 percent poverty threshold creates compound treatment, activating OZ, NMTC, and other LIC-linked programs simultaneously. The null on the full bundle is more powerful than a null on OZs alone.

Third, the RDD identifies effects local to tracts near 20 percent poverty, which may not generalize to very high- or low-poverty tracts.

Fourth, the direct data center location data capture only large facilities reporting to EIA or EPA—smaller colocation facilities and edge data centers are missed. However, these large facilities represent the overwhelming majority of capital investment and are precisely the type that tax incentives aim to attract.

Fifth, the study period (2018–2023) may be too short for data center investment cycles spanning 3–5 years. The dynamic RDD, however, shows no trend toward emerging effects even five years post-designation.

8. Conclusion

Does crossing the LIC eligibility threshold—which triggers Opportunity Zone eligibility alongside other place-based programs—attract data center investment to distressed communities? Using a regression discontinuity design with both direct data center location

data and employment outcomes for approximately 46,000 census tracts, I find no evidence that it does. The null is precisely estimated across both continuity-based and design-based (local randomization) frameworks, robust to alternative bandwidths, donut designs, placebo cutoffs, and direct facility measurement. The vintage analysis separating pre- and post-2018 data centers confirms that OZ designation attracted no new facilities: the null holds for the treatment-relevant flow measure while pre-2018 facilities pass the placebo test.

Read alongside Gargano and Giacoletti (2025) and Jaros (2026), these results reveal a hierarchy of data center incentive effectiveness that carries clear policy implications. State-level tax exemptions that directly reduce operating costs can shift data center location decisions between states—but generate no employment. Local tax abatements that attract facilities create fiscal burdens that persist for decades. And tract-level capital gains incentives that operate through investor intermediaries fail to attract data centers at all. For policymakers in both developed and developing countries, the lesson is that building infrastructure creates the conditions for data center investment; tax incentives—particularly those targeting investors rather than operators—do not substitute for fiber and power.

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

Acknowledgements

References

Ahlfeldt, Gabriel M, Stephen J Redding, Daniel M Sturm, and Nikolaus Wolf, "The Economics of Density: Evidence from the Berlin Wall," *Econometrica*, 2015, *83* (6), 2127–2189.

Bartik, Timothy J, *Who Benefits from State and Local Economic Development Policies?*, Kalamazoo, MI: W.E. Upjohn Institute for Employment Research, 1991.

_ , "Who Benefits from State and Local Economic Development Policies?," *Upjohn Institute Working Paper*, 2019, (19-307).

Busso, Matias, Jesse Gregory, and Patrick Kline, "Assessing the Incidence and Efficiency of a Prominent Place Based Policy," *American Economic Review*, 2013, *103* (2), 897–947.

_ , **John DiNardo, and Justin McCrary**, "New Evidence on the Finite Sample Properties of Propensity Score Reweighting and Matching Estimators," *Review of Economics and Statistics*, 2014, *96* (5), 885–897.

Calonico, Sebastian, Matias D Cattaneo, and Rocío Titiunik, "Robust Nonparametric Confidence Intervals for Regression-Discontinuity Designs," *Econometrica*, 2014, *82* (6), 2295–2326.

Carl Vinson Institute of Government, "Data Center Sales and Use Tax Exemption: Tax Incentive Evaluation," Technical Report, University of Georgia 2025.

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

_ , **Michael Jansson, and Xinwei Ma**, "Simple Local Polynomial Density Estimators," *Journal of the American Statistical Association*, 2020, *115* (531), 1449–1455.

_ , **Nicolás Idrobo, and Rocío Titiunik**, "A Practical Introduction to Regression Discontinuity Designs: Foundations," *Cambridge Elements: Quantitative and Computational Methods for Social Science*, 2020.

_ , _ , **and** _ , *A Practical Introduction to Regression Discontinuity Designs: Foundations* Cambridge Elements: Quantitative and Computational Methods for Social Science, Cambridge University Press, 2020.

Chen, Jiafeng, Edward L Glaeser, and David Wessel, "The Effect of Opportunity Zones on Real Estate Investment: Evidence from Tax Returns," *NBER Working Paper*, 2023, (31004).

Cisco Systems, "Cisco Annual Internet Report, 2018–2023," Technical Report, Cisco Systems 2023.

Data Center Knowledge, "Opportunity Zones: What Data Center Investors Need to Know," *Data Center Knowledge*, 2019.

Duranton, Gilles and Diego Puga, "Micro-Foundations of Urban Agglomeration Economies," *Handbook of Regional and Urban Economics*, 2004, *4*, 2063–2117.

Frandsen, Brigham R, "Party Bias in Union Representation Elections: Testing for Manipulation in the Regression Discontinuity Design When the Running Variable is Discrete," *Advances in Econometrics*, 2017, *38*, 281–315.

Freedman, Matthew, Shanthi Khanna, and David Neumark, "The Impact of Opportunity Zones on Zone Residents," *Journal of Urban Economics*, 2023, *133*, 103407.

Gargano, Antonio and Marco Giacoletti, "Subsidizing the Cloud," Technical Report 5881105, SSRN 2025. Available at SSRN: https://ssrn.com/abstract=5881105.

Gaubert, Cecile, Patrick Kline, and Danny Yagan, "Place-Based Redistribution," *NBER Working Paper*, 2021, (28337).

Gelman, Andrew and Guido Imbens, "Why High-Order Polynomials Should Not Be Used in Regression Discontinuity Designs," *Journal of Business & Economic Statistics*, 2019, *37* (3), 447–456.

Glaeser, Edward L, "The Economics Approach to Cities," *Journal of Economic Perspectives*, 2008, *22* (4), 3–26.

Good Jobs First, "Data Centers: Key Reforms for State Subsidy Legislation," Technical Report, Good Jobs First 2025.

Greenstone, Michael, Richard Hornbeck, and Enrico Moretti, "Identifying Agglomeration Spillovers: Evidence from Winners and Losers of Large Plant Openings," *Journal of Political Economy*, 2010, *118* (3), 536–598.

Hahn, Jinyong, Petra Todd, and Wilbert Van der Klaauw, "Identification and Estimation of Treatment Effects with a Regression-Discontinuity Design," *Econometrica*, 2001, *69* (1), 201–209.

Imbens, Guido W and Thomas Lemieux, "Regression Discontinuity Designs: A Guide to Practice," *Journal of Econometrics*, 2008, *142* (2), 615–635.

International Finance Corporation, "Data Centers in Emerging Markets: Opportunities and Challenges," Technical Report, World Bank Group 2022.

Jaros, Brian J., "Tax Abatements for Data Centers and Government Output," Technical Report 6137769, SSRN 2026. Available at SSRN: https://ssrn.com/abstract=6137769.

JLL Research, "Data Center Outlook: North America," Technical Report, Jones Lang LaSalle 2024.

Kassam, Karim, Christian Krekel, and Nemit Shroff, "Opportunity Zones and Business Entry: Evidence from a Regression Discontinuity Design," *AEA Papers and Proceedings*, 2024, *114*, 436–440.

Kennedy, Patrick and Harrison Wheeler, "Opportunity Zone Investment Activity: An Early Look," Technical Report, National Bureau of Economic Research 2022. NBER Working Paper.

Kline, Patrick and Enrico Moretti, "Place Based Policies with Unemployment," *American Economic Review*, 2013, *103* (3), 238–243.

Kolesár, Michal and Christoph Rothe, "Inference in Regression Discontinuity Designs with a Discrete Running Variable," *American Economic Review*, 2018, *108* (8), 2277–2304.

Lee, David S, "Randomized Experiments from Non-Random Selection in U.S. House Elections," *Journal of Econometrics*, 2008, *142* (2), 675–697.

__ **and David Card**, "Regression Discontinuity Inference with Specification Error," *Journal of Econometrics*, 2008, *142* (2), 655–674.

__ **and Thomas Lemieux**, "Regression Discontinuity Designs in Economics," *Journal of Economic Literature*, 2010, *48* (2), 281–355.

Masanet, Eric, Arman Shehabi, Nuoa Lei, Sarah Smith, and Jonathan Koomey, "Recalibrating Global Data Center Energy-Use Estimates," *Science*, 2020, *367* (6481), 984–986.

McCrary, Justin, "Manipulation of the Running Variable in the Regression Discontinuity Design: A Density Test," *Journal of Econometrics*, 2008, *142* (2), 698–714.

Moretti, Enrico and Daniel J Wilson, "The Effect of State Taxes on the Geographical Location of Top Earners," *American Economic Review*, 2017, *107* (7), 1858–1903.

Neumark, David and Helen Simpson, "Place-Based Policies," in Gilles Duranton, J. Vernon Henderson, and William C. Strange, eds., *Handbook of Regional and Urban Economics*, Vol. 5, Elsevier, 2015, pp. 1197–1287.

Serrato, Juan Carlos Suárez and Owen Zidar, "Who Benefits from State Corporate Tax Cuts? A Local Labor Markets Approach with Heterogeneous Firms," *American Economic Review*, 2016, *106* (9), 2582–2624.

Slattery, Cailin and Owen Zidar, "Evaluating State and Local Business Incentives," *Journal of Economic Perspectives*, 2020, *34* (2), 90–118.

U.S. Census Bureau, "LEHD Origin-Destination Employment Statistics (LODES) Technical Documentation," Technical Report, Center for Economic Studies 2023.

U.S. Energy Information Administration, "Form EIA-860 Detailed Data with Previous Form Data (EIA-860A/860B)," Technical Report, U.S. Department of Energy 2023.

U.S. Government Accountability Office, "Opportunity Zones: Improved Oversight and Data Could Help Ensure Tax Benefit Goes to Low-Income Communities," Technical Report GAO-22-104380, GAO 2022.

World Bank, "Digital Progress and Trends Report 2023," Technical Report, World Bank Group 2023.

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/acs5). 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), CNS09 (information sector). Block-level data aggregated to tracts using first 11 characters of geocode.

Opportunity Zone Designations. Official list of 8,764 designated tracts from the CDFI Fund (https://www.cdfifund.gov/).

EIA Form 860. Plant-level generator data with coordinates and NAICS codes from the U.S. Energy Information Administration (https://www.eia.gov/electricity/data/eia860/) (U.S. Energy Information Administration, 2023). Filtered to NAICS 518* and keyword-matched data center facilities.

EPA GHGRP. Facility-level greenhouse gas reporting data from the EPA (https://ghgdata.epa.gov/ghgp/). Filtered to NAICS 518210 (Data Processing, Hosting, and Related Services).

A.2 Sample Construction

Starting from 72,274 census tracts with non-missing poverty data and positive population:

1. **Poverty RDD sample:** Exclude tracts eligible for OZ only through the MFI pathway.

2. **Employment merge:** Retain tracts with at least one year of LODES employment data.

3. **Data center merge:** Merge geocoded data center locations by census tract FIPS code.

4. **Missing covariates:** Drop tracts with missing values for all baseline covariates.

A.3 Variable Definitions

- **Poverty rate:** Population below poverty / total population $\times$ 100. Source: ACS 2011–2015.

- **OZ designated:** Indicator for CDFI-certified Qualified Opportunity Zone.

- **Pre-period employment:** Average across 2015–2017 from LODES WAC.

- **Post-period employment:** Average across 2019–2023 from LODES WAC.

- **Δ Employment:** Post minus pre average.

- **Data center presence** (`dc_any`): Binary indicator for ≥ 1 data center facility in tract.

- **Data center count** (`dc_count`): Number of data center facilities in tract.

B. Identification Appendix

B.1 McCrary Density Test Details

The McCrary test uses `rddensity` (Cattaneo et al., 2020a). The null is density continuity at the cutoff.

B.2 Covariate Balance Details

For each pre-determined covariate, I estimate a separate RDD using `rdrobust` with MSE-optimal bandwidth. Table 4 reports results.

B.3 Local Randomization Covariate Balance

Table 12 reports Fisher exact p-values from `rdrandinf` for pre-determined covariates within the ± 1.0 percentage-point window around the cutoff. This distribution-free test complements the standard covariate balance checks and is valid even with discrete running variables. As with the standard tests, socioeconomic covariates correlated with poverty show imbalance, while pre-treatment employment levels—the variables most relevant to the identifying assumption—are balanced.

B.4 Donut RDD Details

The donut RDD excludes tracts within a specified distance of the cutoff, addressing measurement error and potential anticipatory behavior at the boundary.
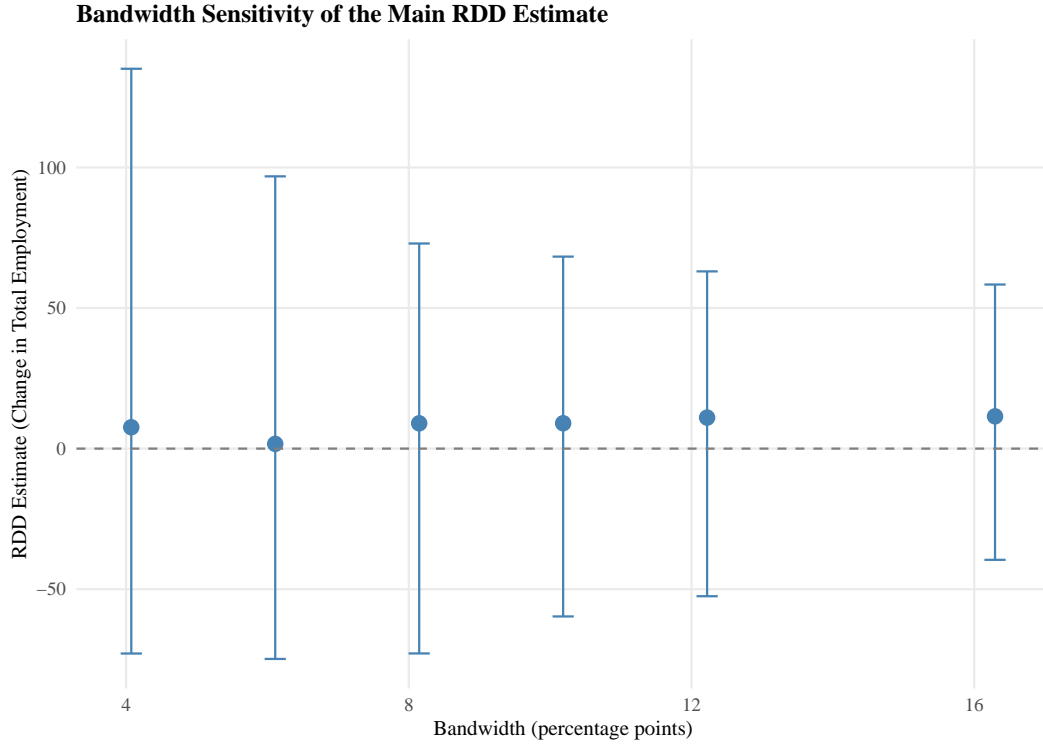
Table 12: Local Randomization Covariate Balance

Covariate	Test Statistic	Fisher p	N
Population	-157.492	0.022	1655
% Bachelor's degree	-1.472	0.000	1655
% White	-4.832	0.000	1655
Median home value	-8913.203	0.131	1642
Unemployment rate	1.014	0.000	1655
Pre-period total employment	32.389	0.888	1655
Pre-period info employment	7.938	0.532	1655

Notes: Fisher randomization inference for covariate balance within ± 1.0 percentage-point window of the 20% poverty cutoff, using `rdrandinf` (Cattaneo et al., 2015) with 1,000 permutations.

C. Robustness Appendix

C.1 Full Bandwidth Sensitivity

Figure 11: Bandwidth Sensitivity of the Main RDD Estimate

Notes: RDD estimates for Δ Total Employment at varying bandwidths. Error bars: robust bias-corrected 95 percent CIs.

C.2 Donut RDD Estimates

Table 13: Donut RDD Estimates

Outcome	Donut	Estimate	Robust SE	p-value	N
Delta Total Emp	±0.5	-26.889	(34.836)	0.327	9,268
Delta Info Emp	±0.5	-2.131	(4.157)	0.747	11,718
Delta Construction Emp	±0.5	-5.462	(4.931)	0.167	8,617
Delta Total Emp	±1.0	-57.313	(49.452)	0.154	7,469
Delta Info Emp	±1.0	-4.092	(7.064)	0.467	7,075
Delta Construction Emp	±1.0	-5.106	(6.115)	0.260	9,073
Delta Total Emp	±2.0	28.950	(47.138)	0.452	13,984
Delta Info Emp	±2.0	-9.689	(8.017)	0.178	10,629
Delta Construction Emp	±2.0	1.591	(6.013)	0.895	13,492

Notes: Donut RDD excludes observations within the specified distance of the 20% cutoff. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

C.3 Polynomial Sensitivity

Table 14: Polynomial Order Sensitivity

Outcome	Poly Order	Estimate	Robust SE	p-value	N
Delta Total Emp	1	8.995	(29.396)	0.818	16,372
Delta Info Emp	1	-1.804	(3.391)	0.724	15,690
Delta Construction Emp	1	-0.497	(6.290)	0.853	15,259
Delta Total Emp	2	1.224	(46.747)	0.961	14,149
Delta Info Emp	2	4.579	(6.156)	0.305	12,468
Delta Construction Emp	2	-1.002	(8.350)	0.946	20,023
Delta Total Emp	3	5.752	(51.182)	0.824	19,755
Delta Info Emp	3	6.241	(6.977)	0.254	19,274
Delta Construction Emp	3	0.086	(10.669)	0.955	21,437

Notes: RDD estimates with varying polynomial orders. All use MSE-optimal bandwidth with triangular kernel. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 15: Kernel Function Sensitivity

Outcome	Kernel	Estimate	Robust SE	p-value	N
Delta Total Emp	triangular	8.995	(29.396)	0.818	16,372
Delta Info Emp	triangular	-1.804	(3.391)	0.724	15,690
Delta Construction Emp	triangular	-0.497	(6.290)	0.853	15,259
Delta Total Emp	uniform	3.163	(35.377)	0.953	8,160
Delta Info Emp	uniform	2.601	(5.095)	0.500	7,061
Delta Construction Emp	uniform	-0.789	(5.838)	0.747	11,715
Delta Total Emp	epanechnikov	5.760	(30.526)	0.950	13,331
Delta Info Emp	epanechnikov	-2.305	(3.361)	0.584	14,940
Delta Construction Emp	epanechnikov	-0.598	(5.908)	0.812	14,685

Notes: RDD estimates with alternative kernel functions. All use MSE-optimal bandwidth and local linear regression. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 16: Inference Robustness

	(1) Baseline rdrobust	(2) + Covariates rdrobust	(3) Parametric HC1	(4) County Clustered
Δ Total Emp	8.995	10.603	6.536	6.536
	(29.396)	(29.682)	(28.382)	(31.333)
Δ Info Emp	-1.804	-1.356	-0.528	-0.528
	(3.391)	(3.432)	(3.327)	(3.228)
Δ Construction Emp	-0.497	-0.356	0.525	0.525
	(6.290)	(6.558)	(5.707)	(6.163)

Notes: Column (1): baseline `rdrobust` with MSE-optimal bandwidth and triangular kernel. Column (2): `rdrobust` with demographic covariates (population, education, race, unemployment). Column (3): parametric linear RDD within optimal bandwidth with HC1 standard errors. Column (4): parametric linear RDD with county-clustered standard errors. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 17: Local Randomization Inference

Outcome	Window (pp)	Test Stat	Fisher p	N_{left}	N_{right}
Δ Total Emp	± 0.50	38.457	0.561	321	497
Δ Info Emp	± 0.50	2.958	0.750	321	497
Δ Total Emp	± 0.75	43.009	0.338	491	736
Δ Info Emp	± 0.75	5.694	0.283	491	736
Δ Total Emp	± 1.00	19.965	0.575	670	985
Δ Info Emp	± 1.00	3.092	0.433	670	985

Notes: Randomization inference using `rdrandinf` (Cattaneo et al., 2015) with 1,000 permutations. Windows are symmetric around the 20% poverty cutoff. Fisher p-values test the sharp null of no treatment effect for any unit within the window.

Table 18: Data Center Presence: Bandwidth and Donut Robustness

Test	Specification	Estimate	Robust SE	p-value	N
Bandwidth	1.00× optimal	-0.00017	(0.00022)	0.296	5,178
Bandwidth	1.50× optimal	-0.00014	(0.00021)	0.297	7,940
Bandwidth	2.00× optimal	0.00011	(0.00031)	0.190	10,904
Donut	±0.5 pp	0.00000	(0.00016)	0.161	2,832
Donut	±1.0 pp	-0.00063	(0.00065)	0.421	5,421
Donut	±2.0 pp	0.00222	(0.00206)	0.215	7,216

Notes: Nonparametric RDD estimates for binary data center presence using `rdrobust` with triangular kernel. Bandwidth rows vary the bandwidth as a fraction of the MSE-optimal bandwidth. Donut rows exclude tracts within the specified distance of the 20% poverty cutoff. Robust bias-corrected inference throughout. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.
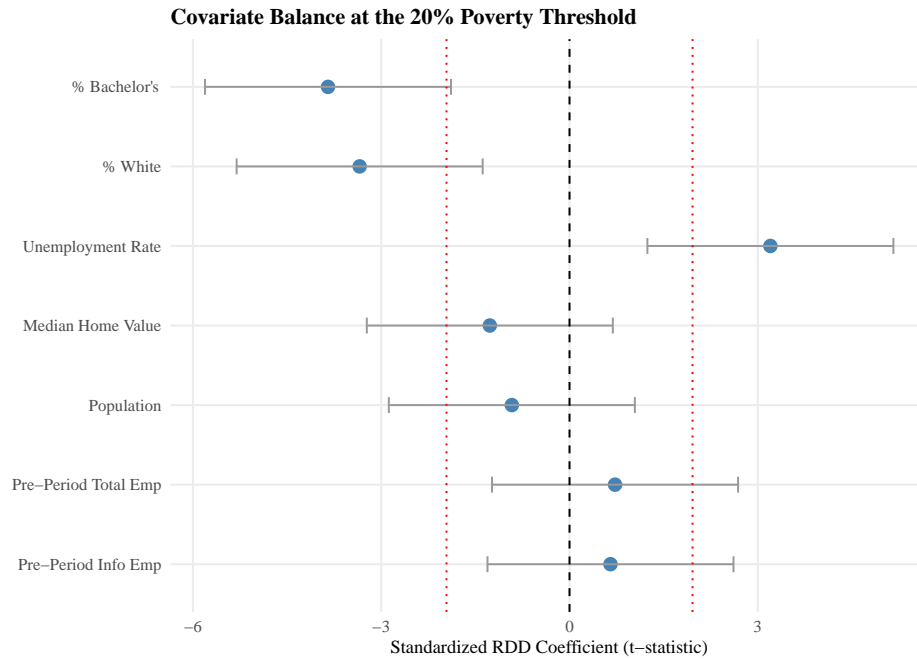
C.4 Kernel Sensitivity

C.5 Inference Robustness

C.6 Local Randomization Inference

C.7 Data Center Robustness

C.8 Covariate Balance Figure

Covariate Balance at the 20% Poverty Threshold

Figure 12: 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 (±1.96).

D. Heterogeneity Appendix

The urban/rural heterogeneity analysis defines "urban" tracts as those with total population exceeding 2,000, approximately corresponding to the Census Bureau's urban area definition.