The Local Economic Impact of Data Centers

Dany Bahar[†] Greg C. Wright[‡]

[†]Brown University. dany_bahar@brown.edu
[‡]University of California, Merced. gwright4@ucmerced.edu

July 7, 2026

Abstract

A hyperscale data center costs over a billion dollars but employs only dozens of people. We ask whether these investments generate the local agglomeration spillovers that large plant openings have historically produced. Comparing built facilities to fifty-two announced-then-cancelled ones, we find a large but sharply localized capital footprint in nighttime lights that fades within five kilometers, and no evidence of agglomeration spillovers. Local wages, supplier clustering, and downstream industry growth show no response, on estimates precise enough to rule out meaningful gains. Hiring and new-firm formation show no increase, though the hiring estimate is less precise. The local incidence falls on land and local government budgets, as rents rise by a few percent without net in-migration, school revenues and public-works spending increase, and retail electricity rates fall rather than rise. Data centers are agglomeration-scale capital without agglomeration.

Keywords: data centers; place-based investment; local labor markets; nighttime lights; agglomeration.

JEL Codes: R11, R23, O33, H71.

1 Introduction

The physical infrastructure of digital services has become a major component of US capital investment. Data centers consumed roughly four percent of US electricity in 2023, and the rise of artificial intelligence has sharply accelerated new construction. A single hyperscale campus typically costs five hundred million to two billion dollars and draws one hundred to five hundred megawatts of continuous power, yet employs only thirty to fifty people on site. At least thirty-five states now offer tax incentives aimed specifically at data centers, with cumulative subsidies approaching twenty billion dollars.[1] Those subsidies are justified by the promise of local economic gains. This paper asks whether the gains materialize.

The natural benchmark is the canonical result on large plant openings. Greenstone et al. [2010] identify these spillovers by comparing counties that won a large new plant to the runner-up counties that lost it, and show that the productivity of incumbent establishments in winning counties rises by about twelve percent over the following five years. The gain operates through three agglomeration channels, namely a deeper shared labor pool, denser supplier linkages, and knowledge spillovers among nearby firms. This result is the empirical backbone of the case for place-based subsidies. At the same time, subsequent work contests this result, finding that estimates from the winner-versus-loser design are sensitive to the choice of comparison counties [Patrick, 2016], and the average large plant shows little evidence of productivity spillovers [Patrick and Partridge, 2022]. The spillover evidence is therefore contested even for manufacturing plants, the historically strongest case for agglomeration.

A data center is large in terms of investment dollars, but it employs dozens of people, not thousands, and so does not deepen the local labor pool. Moreover, its main inputs (electricity and fiber-optic connectivity) are purchased through national markets and its output (compute) is delivered over the cloud to customers around the world. Given these structural differences from manufacturing plants, it is an open question whether the canonical agglomeration result extends to the signature capital investment of the AI era. The subsidies at stake make the question a consequential one.

Operators screen sites on cheap power, grid capacity, land cost, and fiber, so a county that hosts a data center differs from one that does not in ways that also shape the local economy. We document that this selection cannot be easily instrumented away. All of the predetermined features that we test as predictors of hyperscale siting are also correlated with local industrial character, and the features that are clean do not predict where these facilities go (Section 3).

Our primary design is therefore the analog of the winner-versus-loser comparison Greenstone et al. [2010] used to identify plant spillovers. We compare built facilities to announced-then-cancelled ones, which were sited for the same reasons but never built. The cancelled cohort contains eighty-four candidate projects, of which fifty-two enter the headline design. We also report two supporting designs. First, we apply a within-site ring comparison, comparing each facility's inner

[1] Authors' tally from the Good Jobs First Subsidy Tracker, which records state and local data-center subsidy awards and is one of the sources for our facility registry (Section 2).

rings to its own twenty-five to fifty kilometer ring, which provides estimates for outcomes that pass a pre-trend test. Second, we apply a county comparison design that estimates more aggregate employment effects. Using satellite-detected land clearing, daytime imagery, and regulatory filings we observe, separately for every facility, the date construction begins and the date operations begin, so the build and the operation enter the event studies as distinct treatments rather than a single opening date.

We first establish that the capital footprint is real, large, and local. Relative to cancelled sites, nighttime lights at built sites jump at the start of construction within the first kilometer and decay to near zero by five kilometers, and the ring design recovers the same magnitude with a flat pre-trend. Daytime imagery confirms the timing, showing that vegetation at the parcel is cleared exactly when the lights turn on. This observed footprint is useful because it fixes the scale of the treatment. We then turn to the evidence for agglomeration, testing each of the three mechanisms. Hiring in the surrounding rings shows no detectable rise, though that estimate is imprecise. Local wages do not rise, on a panel large enough to make the null informative, and workplace employment shows no jump at opening. The sectors that build and equip a data center do not cluster near the site. The industries that use data centers do not subsequently form a local cluster, and new firm formation shows no jump in business-birth records. We confirm this in the realized county employment data, where the positive effect is confined to the data-center sector itself and its network, with no growth in any downstream or compute-using industry. Colocation data centers, which add capital but require little new construction, serve as a within-design placebo, showing no capital footprint and no spillover through any channel.

What local incidence we find falls on land, not labor. Residential rents rise modestly in the facility's catchment, with no multifamily permitting response, so the increase capitalizes into the existing housing stock rather than reflecting a demand boom. Net migration is near zero, because gross in-migration rises but is offset by comparable out-migration. Local public finances strengthen, as school revenues and public-works spending rise, so part of the local gain accrues to municipal balance sheets alongside landowners. Foot traffic in the surrounding rings shows no robust change relative to cancelled sites, and commercial spending shows no robust response. Retail electricity rates, the most common objection to hosting a facility, fall rather than rise after a hyperscale opening, because the new load spreads the fixed costs of the local distribution network. Electricity rates are unaffected by colocation openings.

Taken together, the results describe capital without agglomeration. Data centers have recently been studied at the county level [Bahar and Wright, 2026, Alvarez et al., 2026].[2] Our within-site design adds the sub-county capital footprint, direct tests of each agglomeration channel, and the incidence on land. Our contribution is threefold. First, we provide a sub-county test of whether agglomeration spillovers extend to the capital frontier of the AI era, and find that they do not.

[2]Our own county-level analysis [Bahar and Wright, 2026] reported a positive downstream, compute-using association using a synthetic-control design. As Section 5.5 shows, that association does not survive a staggered-robust estimator and is unaccompanied by any establishment entry, so we read it as a pre-existing trend rather than a local agglomeration spillover.

Second, we show that this is because each agglomeration channel is absent following the investment. Third, we trace the incidence to land and capital rather than labor, which bears on the design of the subsidies now flowing to these facilities. More broadly, the result defines a limit on the local-multiplier logic of industrial policy as investment shifts from labor-intensive plants toward labor-light digital capital.

The remainder of the paper proceeds as follows. Section 2 describes the data. Section 3 sets out the identification strategy. Section 4 documents the capital footprint. Section 5 tests the agglomeration channels and the county employment composition. Section 6 traces the local incidence to land. Section 7 examines retail electricity rates. Section 8 interprets the findings. Section 9 concludes. Appendices report the instrument search, the geography of compute, and supplementary results.

2 Data

Data center registry. We build a master registry of US data centers from four sources. The base layer is the Open Source Data Center Atlas (Pacific Northwest National Laboratory's IM3 project), which identifies 1,242 facilities from OpenStreetMap footprints tagged as data centers, with coordinates and, where available, square footage and operator. We augment it with operational-date information from operator press releases, state economic-development announcements, the Good Jobs First subsidy tracker, and trade press, and with four research batches targeting state announcements, specialty wholesale operators, AI-wave operators, and REIT filings. The final registry contains 1,494 facilities. Of these, 341 are hyperscale-tier, and 339 of the 341 have operational dates inside the 2012–2024 VIIRS window. We assign hyperscale tier by operator identity (Amazon, Microsoft, Google, Meta, Apple, Oracle, plus hand-curated large specialty and AI-wave campuses). We do not impose a megawatt threshold, since capacity is not consistently observed. We also assemble the cancelled cohort for the built-versus-cancelled primary design. It contains eighty-four announced-then-cancelled projects, of which fifty-two have the coordinates and satellite coverage to enter the nighttime-lights design. Section 3 reports the cohort's sponsor composition and shows that the capital result is robust to restricting the control arm to projects with a named hyperscale tenant.

Nighttime lights. We extract monthly mean radiance from the Visible Infrared Imaging Radiometer Suite (VIIRS) Day/Night Band Monthly Composite for April 2012 through December 2024, in concentric rings of 0–1, 1–2, 2–5, 5–10, 10–25, and 25–50 kilometers around each facility, aggregated to annual means. We classify each facility as isolated if no other data center within two kilometers opened earlier, and use the isolated, construction-dated sample for the headline capital-footprint estimate.

Local labor, housing, and spending panels. We measure local labor demand from the LinkUp Job Records panel [LinkUp, 2025] (postings by employer and geocoded location, by ring and quarter), residential rents from the RentHub Rental Data panel [RentHub, 2022] (median list price by ring and month), consumer spending from the SafeGraph Spend Patterns panel [SafeGraph, 2022], and foot traffic from the Advan Weekly Patterns panel [Advan Research, 2025], all distributed by Dewey Data. Each enters the within-site ring design dated to operations, with state-by-month fixed effects.

Realized county employment. For the county-level composition analysis we use industry employment and establishment counts from the Quarterly Census of Employment and Wages, which align with the standard County Business Patterns employment and establishment measures. We focus on data processing (NAICS 518), telecommunications (517), web and information services (519), IT consulting (5415), and professional, scientific, and technical services (541), with manufacturing (31–33) and population-serving sectors as placebos.

Instrument-search inputs. The instrument search draws on predetermined county features, namely 345 kV transmission-substation density, the InterTubes long-haul fiber network [Durairajan et al., 2015], the 1980 county share of urban college population (NHGIS), station-level climate normals (CustomWeather), and the EIA-860 generation fleet (Appendix A).

2.1 Dating construction and operations

For every facility we observe two dates rather than one. The operational opening date comes from the registry. The construction start date is recovered from daytime satellite imagery. In this case, a hyperscale build begins with land clearing, which appears as a sustained drop in the parcel's vegetation index combined with a rise in its built-up surface index in the Landsat and Sentinel-2 scenes. Vegetation cover at the parcel is flat until the construction year and falls sharply after it. As an independent check, we join Federal Aviation Administration obstruction-evaluation filings to our projects and these recover essentially the same dates (Section 3.5). Observing the build and the opening separately lets the event studies treat construction and operations as distinct treatments.

3 Research Design

Data centers do not locate at random. Operators screen on cheap power, available grid capacity, large parcels of low-cost land, fiber proximity, and tax policy, so a county that hosts a data center differs from one that does not in ways that also move the local economy. A cross-county comparison would confound the data center with the conditions that attracted it. We document this selection directly. We attempted to instrument data center siting with a range of plausibly exogenous predetermined features, and each instrument failed, for reasons we report in Appendix A. The features that predict siting (power, industrial land, fiber) are themselves correlated with local industrial character, and the features that are clean (climate, historical human capital) do not predict where hyperscale facilities go.

We instead follow the design that anchors the agglomeration literature. Greenstone et al. [2010] identify plant spillovers by comparing counties that won a large plant to the runner-up counties that lost it, on the logic that winner and loser were selected on the same fundamentals. Our primary design is the data-center analog where we compare built facilities to announced-then-cancelled ones. Around it we use two supporting designs, each for what it does best. The within-site ring comparison supplies precision and spatial detail for outcomes that pass its pre-trend test, and a county comparison recovers the sectoral composition of aggregate employment effects.

3.1 Built versus cancelled facilities

A facility that was announced and sited but never built is the loser to a built facility's winner. Built and cancelled sites were chosen for the same grid, land, fiber, and tax fundamentals, so a post-opening difference between them primarily reflects the fact that the facility was built, rather than the conditions that attracted it. We assemble eighty-four announced-then-cancelled hyperscale-scale projects from trade-press coverage, regulatory filings, and operator announcements, of which fifty-two enter the nighttime-lights design, and estimate, for each outcome,

$$y_{i,r,t} = \sum_k \left[\beta_k \mathbf{1}[t - t_i^c = k] \times \text{Near}_{i,r} \times \text{Built}_i + \delta_k \mathbf{1}[t - t_i^c = k] \times \text{Near}_{i,r}\right] + \alpha_{i,r} + \lambda_t + u_{i,r,t}, \quad (1)$$

where Built_i marks facilities actually built and t_i^c is the construction-start (for cancelled sites, scheduled) date. The path of β_k traces what building the facility does to its surroundings, net of anything common to places selected for one. For the capital footprint the comparison is especially sharp since a parcel that is never built never lights up, whatever the reason for cancellation.

The composition of the cancelled cohort deserves explicit note. Of the eighty-four candidate projects we track, fourteen had a named hyperscale operator publicly attached at cancellation, nineteen were sponsored by established data-center operators without a named tenant, and the remaining fifty-one, about sixty percent, were speculative developments by real-estate sponsors. A speculative loser is a weaker analog to the runner-up counties in Greenstone et al. [2010] than an operator-committed one, since it carries no signed tenant or confirmed grid allocation. Two facts limit the concern for our use of the design. For the capital footprint the comparison is mechanical, since a parcel that is never built never lights up whatever the sponsor. And the result does not depend on the speculative majority. Restricting the cancelled arm to the nine isolated sites where a hyperscale operator was named leaves the groundbreaking-year jump essentially unchanged, at seventy percent versus sixty-eight percent in the full isolated arm, though so thin a subset cannot support longer horizons. The cancelled sample is also somewhat small overall, and most cancellations followed local opposition, so cancelled sites may sit in somewhat less development-friendly places.[3] This is in part why we pair the design with the better-powered ring comparison

[3] Because most cancelled projects were announced late in the sample, the cancelled arm contributes few post-event observations. Coverage at event years zero through four is 51, 23, 6, 4, and 1 sites, so the built-versus-cancelled path at longer horizons is identified primarily from the built arm's post-opening change against the common pre-event trend rather than from post-period movement in the cancelled arm. Section 4 quotes magnitudes only through the

below. In addition, built sites are dated by satellite-detected land clearing, while cancelled sites carry scheduled dates from filings and trade press, which are softer. However, the independent regulatory-filing dates described in Section 3.5 validate the built-site dates as well.

3.2 Within-site rings

The cancelled comparison establishes whether an effect exists. It is too small to resolve how an effect decays over space or to deliver tight confidence intervals on nulls. For those we use the within-site ring design, which compares an inner ring of a facility to the same facility's twenty-five to fifty kilometer outer ring,

$$\log(1 + \mathrm{NTL}_{i,r,t}) = \sum_k \beta_k \mathbf{1}[t - t_i^c = k] \times \mathrm{Near}_{i,r} + \alpha_{i,r} + \gamma_{r,t} + \varepsilon_{i,r,t}, \tag{2}$$

where i indexes the facility, r the ring, t the year, t_i^c the facility's construction-start year, $\mathrm{Near}_{i,r}$ marks the inner ring, $\alpha_{i,r}$ is a facility-by-ring fixed effect, and $\gamma_{r,t}$ is a ring-by-year fixed effect. The facility-by-ring fixed effect absorbs any time-invariant difference between a site's inner and outer ring, including the operator's siting preferences, the local industrial structure, and grid and fiber access. Identification comes from the change in the inner ring relative to the outer ring of the same site.

Two features of the design matter for credibility, and both correct a pre-trend present in a naive version of this specification. First, we date the event to construction start, not to the operational opening, to avoid confusing pre-period construction activity with operations (Section 2.1). Dated to construction, the inner-ring brightness is flat before groundbreaking and jumps at the groundbreaking date.

Second, we restrict the headline to isolated sites, meaning facilities with no other data center within two kilometers that opened earlier. A new facility built beside an existing one inherits the neighbor's brightness in its inner ring before it breaks ground, which could also generate a false pre-trend. The isolated, construction-dated event study has a flat pre-trend and then a jump of thirty-eight percent in the construction year that builds to a larger footprint over the following years. The jump survives controlling for a site-specific linear pre-trend estimated across the pre-construction years (the pre-trend is flat).

Importantly, we only report the ring estimates for outcomes whose pre-construction leads are flat. The capital footprint, rents, advertised wages, and firm entry pass this test. Posting counts and workplace employment do not, so for those outcomes the cancelled-project design carries our interpretation and we do not lean on the ring estimates in either direction. So that the reader can audit this routing rather than take it on faith, Appendix Table 8 reports every outcome under every design in which it can be estimated, with the estimate, its precision, and the reason any design is excluded for that outcome.

year after groundbreaking for this reason.

3.3 Demand-side outcomes

We use the same within-site ring comparison in equation (2) to identify three demand-side outcomes, namely residential rents, consumer spending, and foot traffic. We date these to the operational opening rather than to the construction start. The rent, spending, and foot-traffic panels are monthly and span the 2020 to 2023 period, so we replace the ring-by-year fixed effect with a state-by-month fixed effect, which absorbs the differential work-from-home and urban-recovery shocks across metropolitan areas in those years. The primary design identifies each of these outcomes, and the rings refine the spatial pattern. Relative to cancelled sites, rents rise modestly at built sites, foot traffic shows no robust change, and commercial spending is null in both arms. The rent, spending, and foot-traffic estimates enter the incidence analysis in Section 6.

3.4 Employment

Local labor demand is harder to identify locally relative to the capital footprint, so we measure it two ways. The first approach applies the within-site ring design of equation (2) to job postings near each facility, counting openings advertised within each ring.

The second approach is at the county level, because a data center's own employees and any firms that cluster around it need not fall inside a single ring. We compare counties that gain a hyperscale data center to counties that do not. The event study exploits County Business Patterns and Quarterly Census of Employment and Wages employment by industry:

$$\log(1 + \mathrm{Emp}^s_{c,t}) = \sum_k \theta^s_k \, \mathbf{1}[t - t^o_c = k] \times \mathrm{DC}_c + \mu_c + \lambda_t + u_{c,t}, \tag{3}$$

for industry s, county c, and the county's first data-center year t^o_c, with county and year fixed effects and never-treated counties as the comparison group. Pre-period leads for the direct sector are jointly null ($p = 0.53$). The leads in some downstream sectors are noisier so, as a result, we read the design through the sectoral composition of the effect, which is what distinguishes a direct effect in the data center's own industry from an agglomeration spillover to surrounding firms. We then corroborate the estimate with a staggered-robust estimator that uses not-yet-treated counties as controls. We report the postings and county employment results in Section 5.

3.5 Credibility of the capital-footprint design

The capital-footprint estimate is robust to the usual checks on a ring design. Dropping each facility in turn leaves the estimate in a narrow band, so no single site drives it, and a donut specification that uses the one-to-two kilometer ring as the inner contrast, omitting the campus pixel, returns the same spatial decay. Because most hyperscale sites sit near one another, we cluster standard errors at the campus, rather than the facility, level which widens the interval but leaves the effect statistically significant. Most important for the construction-dating, we re-date the facilities from Federal Aviation Administration (FAA) obstruction-evaluation filings, a regulatory timestamp recorded years before any radiance response and joined to a facility by location, with no satellite input. The construction-dated capital footprint is essentially unchanged under these independent dates, which establishes that it is not an artifact of nighttime-lights-derived event timing.

In Section 6 we perform one further check with respect to the rent result. Rents rose generally over 2020 to 2023 in the kinds of places that host data centers, and a within-site comparison would attribute that growth to the facility. We therefore build a synthetic control for each facility from population-matched locations with no data center, which nets out the common growth and isolates the data-center-specific rise.

4 The capital footprint

This section describes the capital footprint associated with data centers.

4.1 Built versus cancelled site design

Figure 1 reports the estimates from equation (1). This research design compares the inner-ring (zero to one kilometer) nighttime-lights path at built hyperscale sites to announced-then-cancelled sites, dated to the construction start (for cancelled sites, the scheduled start). Brightness at built sites jumps by roughly sixty percent in the construction year and sits nearly one hundred eighty percent above the cancelled-site path a year later, continuing to build over the following years as campuses phase in additional buildings. A parcel that is never built never lights up, so the contrast isolates the act of building from the siting that both groups share. The footprint derives from activity at the parcel including its security and parking lighting, its substations and cooling plant, and its continuous overnight operation.

Built sites brighten modestly relative to cancelled ones in the years before groundbreaking, which reflects two things. Some built facilities rise beside an earlier neighbor whose light contaminates the inner ring, and early site work precedes the satellite-detected clearing date at some campuses. To remove this contamination we focus on *isolated* sites, which are the forty-eight built facilities with no earlier data center within two kilometers of the parcel and the forty-nine cancelled sites away from operating campuses. For this sample the leads flatten and are jointly indistinguishable from zero ($p = 0.14$ clustered on facility, $p = 0.28$ clustered on campus). The jump at groundbreaking is sixty-eight percent, and the path reaches roughly two hundred thirty percent above the cancelled-site benchmark a year later.

Two properties of this estimate matter for how it should be read. First, the cancelled cohort is concentrated in recent announcement years, so its coverage declines with the event horizon. Forty-eight control sites identify the groundbreaking-year contrast and twenty-one the year-one contrast, but five or fewer identify the contrasts beyond that. We therefore quote magnitudes through the year after groundbreaking, and we read the continued build-up at longer horizons, which reaches several hundred percent by year four, as suggestive rather than precisely identified. Second, the footprint belongs to the campus, not to a single building. Forty of the forty-eight isolated built sites add a sibling building within two kilometers during the event window, so the multi-year build-up reflects campuses phasing in additional buildings. Restricting to the twelve sites that stay single-building through year four, which are also the smallest projects, gives a footprint of about seven percent at groundbreaking, imprecisely estimated. We report the campus-level number as the headline because the campus is what a jurisdiction recruits. Clustering standard errors at the campus level widens the intervals by roughly sixty percent and leaves every post-groundbreaking coefficient statistically significant.

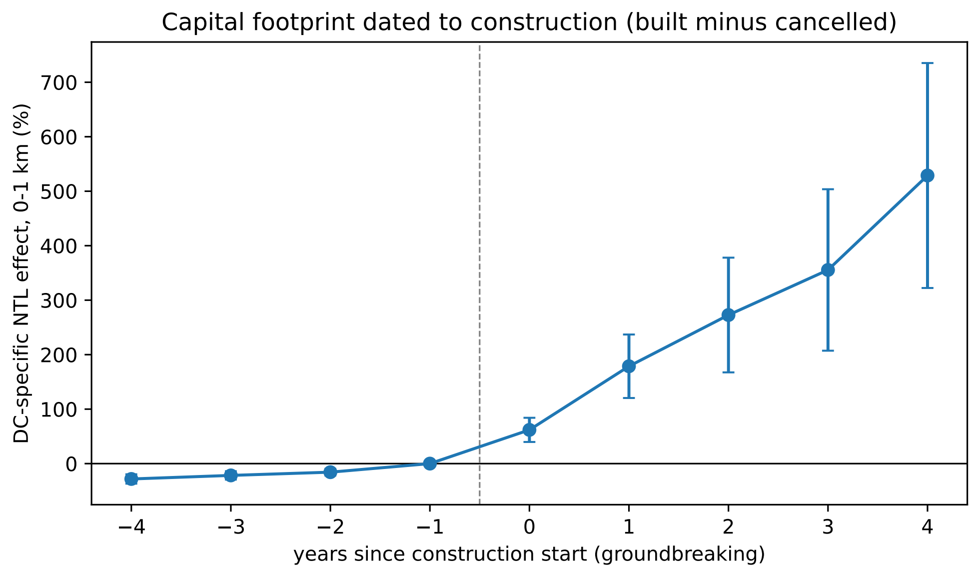

Figure 1: Capital footprint at built versus cancelled sites, dated to construction start

Notes: Event study of equation (1): the inner-ring (0–1 km) log nighttime-lights path at built hyperscale facilities relative to announced-then-cancelled facilities, by year since construction start (scheduled start for cancelled sites), omitting the year before. Built facilities are dated by satellite-detected land clearing. Standard errors clustered on facility. Cancelled-site coverage declines with the event horizon (51, 23, 6, 4, and 1 sites at years zero through four), so estimates beyond year one rest on few control sites and should be read as suggestive.

4.2 Within-site ring design

The within-site ring design sharpens the estimate in two ways, each of which removes a source of contamination in the built-versus-cancelled design. The first is the dating. A hyperscale campus takes two to three years to build, so dating the event to the operational opening places the construction phase, when the parcel is physically changing, before the event window and reads it as a pre-trend. We date each facility instead to the construction-start year, recovered from satellite-detected land clearing (Section 2.1). Observing the construction start and the operational opening separately, for every facility, is also what lets the phase decomposition in Appendix C apportion the footprint between the build and the operation.

The second issue is the sample. Roughly two in five hyperscale sites sit within two kilometers of an earlier data center, such that the site inherits its neighbor's brightness in its inner ring before it breaks ground. Splitting the sample makes this explicit. At sites with an earlier neighbor within two kilometers, inner-ring brightness rises about fifty percent in the years before construction. At isolated sites, where the inner ring is genuinely just the future parcel, the pre-construction excess is seven percent, and the event study is flat before groundbreaking. A joint test of the pre-construction coefficients does not reject ($p = 0.21$), and the jump is thirty-eight percent in the construction year, surviving a control for a site-specific linear trend estimated from the pre-construction years.

The two designs agree on timing and shape. Both show a flat pre-period at isolated sites followed by a sharp jump in the construction year and a multi-year build-up. The point magnitudes are not directly comparable because the two designs transform radiance differently, but each implies a construction-year brightening of several tens of percent at the parcel.

4.3 Spatial decay

The footprint is sharply localized. Estimated ring by ring, the effect is largest at zero to one kilometer, falls by more than half at one to two kilometers, and is small and fading by two to five kilometers, with no detectable effect beyond five kilometers. A donut specification that drops the zero-to-one kilometer ring and uses one-to-two as the inner contrast returns the same decay, so the result is not an artifact of the campus site itself. Overall, the site impact is large but it does not reach far.

4.4 Colocation as a within-design placebo

Colocation facilities add computing capacity inside existing buildings and require little new construction. These show no comparable changes in the capital footprint. Given this, the rest of the paper takes hyperscale facilities as the treatment whose spillovers, or absence of them, we measure.

5 Test of agglomeration

Section 4 establishes that a hyperscale data center is a large, spatially concentrated investment. The question this section asks is whether that capital generates the local spillovers that large plant openings have historically produced. Greenstone et al. [2010] trace those spillovers to three channels, namely a deeper shared labor pool, denser supplier linkages, and knowledge spillovers among nearby firms. We test each channel directly at the facility level, using the within-site ring design for the outcomes that pass its pre-trend test, and the built-versus-cancelled comparison for job postings, which do not. Colocation facilities serve as a within-design placebo here again, with the caveat that they sit in denser markets than hyperscale campuses, so the placebo should not be read as a perfectly matched comparison. We find that none of the three channels operates, and the productivity outcome they would produce does not appear either.

5.1 Labor pooling

Local labor pool depth is the most-studied agglomeration channel and should lead to a rise in hiring and wages in the surrounding rings. In this case, we only apply the built-versus-cancelled design due to the fact that the within-site postings series carries a differential inner-versus-outer trend that the ring comparison cannot separate from an effect. Relative to cancelled sites, postings near built facilities move by +9% ($p = 0.46$) with a flat pre-trend ($p = 0.85$), and the estimate remains statistically indistinguishable from zero across alternative estimators and event timings, including Poisson counts, a long difference that controls for pre-period volume, and construction dating. We are explicit about what this null does and does not establish. The design is imprecise for this outcome. At conventional power it could only detect a hiring increase of roughly forty percent or more, so the postings evidence shows no detectable rise rather than a tight zero, and on its own it cannot rule out moderate positive effects. The sharper evidence on the labor channel comes from the wage and workplace-employment results below, which are well powered. The within-site estimates from two independent job postings vendors (LinkUp and WageScape [WageScape, 2022]) are consistent with zero at every ring but inherit the trend, so we do not lean on them in either direction.

The mean log advertised salary of nearby postings is not affected by hyperscale entry at any ring. The estimate at 0–5 km is -1.3% ($p = 0.65$), measured over facility-by-ring-month cells that average several hundred salaried postings, so the null is well powered. Its ninety-five percent interval, $[-6.9, +4.6]$ percent, rules out all but modest local wage gains. One caveat is that advertised-salary panels are concentrated in salaried, posted positions, so the null is most relevant for white-collar workers and may not capture construction-trade compensation. The salary-disclosure rate near the facility is itself unchanged at opening ($p = 0.43$ at 0–5 km), so the null is not an artifact of differential disclosure. Workplace employment from the Longitudinal Employer-Household Dynamics (LODES) data tells a similar story. The raw within-site hyperscale effect is +8% at 0–5 km, but it does not survive detrending, collapsing to -0.3% with a pre-existing inner-versus-outer slope significant at $p < 0.001$.

Table 1: Job postings ring DiD: divergence from the nighttime-lights spillover

	DC-title	Non-DC-title	All postings
0–1 km	†	+5.9%*	+3.2%
1–2 km	†	+5.8%	+1.9%
2–5 km	-14.3%	+8.0%	+7.7%
Facilities in regression	298 hyperscale-operator postings DCs		

Notes: Poisson Pseudo-Maximum-Likelihood (PPML) estimates of within-site ring DiD on quarterly posting counts, aligned to curated operational dates. DC-title postings are those containing data-center-related keywords. Sample is hyperscale-operator-tagged postings (AWS, Microsoft, Google, Meta, Apple, Oracle, plus close variants). Pooled across archetypes. %lift reported as $(\exp(\hat{\beta}) - 1) \times 100$. † indicates a cell where the post-period count is essentially zero, producing PPML separation rather than an identified estimate. SEs clustered on DC. Stars: $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$.

5.2 Supplier linkages

The second channel is local supplier linkages. A large opening that bought inputs locally would pull supplier activity toward the site. We test this on the postings panel by isolating the sectors that build and equip a data center, namely construction, specialty trades such as electrical and mechanical contracting, electrical-equipment wholesale, architectural and engineering services, and machinery manufacturing. We aggregate postings in these sectors to facility-by-ring-month counts and run the within-site DiD on the composite and on each sector separately.

Supplier postings do not rise near hyperscale or colocation facilities. The hyperscale supplier composite is -1.9% at 0–5 km ($p = 0.87$), and every individual sector is null. Because supplier linkages may operate at the metropolitan level rather than a more local scale, we also run a broad specification that compares the pooled 0–25 km near zone to a 50–150 km far control, which would catch a region-wide supplier rise that the within-site control might absorb. It is null as well, at $+3.6\%$ ($p = 0.73$) for hyperscale. Colocation is null under both designs. The sectors that build a data center do not localize around it, consistent with build crews that travel and equipment that ships in from national markets.

5.3 Knowledge spillovers and firm entry

The third channel is knowledge spillovers among co-located firms. Two implications are testable. The industries that use data centers should cluster near them, and new firms should form in the catchment after a facility opens. We find that the downstream industry does not form a local cluster. Appendix B maps the firms that are the heaviest users of data-center capacity and shows, descriptively, that while they sit somewhat closer to data centers than the average firm, the proximity reflects shared metropolitan location rather than anything the facility causes. Most large compute users are nowhere near a data center, the firms that are nearby pre-date the facility, and exurban data centers sit in compute deserts. The cloud decouples where compute is produced from where it is used, so the knowledge cluster that co-location would require never forms around the site.

New firm formation also shows no increase at opening. We test it on true business births using Secretary-of-State new-registration records for Virginia and Texas geocoded to the facility rings. The within-site jump in new registrations at the immediate 0–5 km ring is +0.5% ($p = 0.42$) in the pooled sample and −0.7% ($p = 0.46$) restricted to hyperscale facilities, the same universe the wage test uses. The pooled estimate is not a uniform zero, and we do not present it as one. As the appendix reports, registrations fall three to four percent near facilities in Virginia's dense data-center corridor and rise roughly ten percent in Iowa, whose active-only registry partly reflects survivorship, so the state-level responses are heterogeneous and offsetting. What matters for the knowledge-spillover channel is that no sample shows the systematic entry increase the channel predicts. Where entry moves at all near hyperscale sites, it moves down. A national county-level design on the Census Business Formation Statistics returns the same null (+1.4%, $p = 0.49$). A placebo that moves the opening date thirty months earlier produces a null jump at every ring, which confirms that the design is not manufacturing the result from a pre-existing trend. The naive estimate at 0–5 km is negative and significant, so the detrending is doing real work. It decomposes the raw drop into no jump at opening plus a pre-existing decline, and we read entry near data centers as drifting down before and after the facility arrives rather than responding to it. New firms do not enter the catchment when a data center opens.

5.4 Incumbent productivity

In Greenstone et al. [2010] incumbent establishments see total factor productivity gains of around twelve percent. We do not measure plant-level productivity directly but instead bound it through its observable effects. Specifically, productivity gains for incumbents will show up either as higher wages or as faster growth and hiring at incumbent firms. Both of these channels are flat.

The wage outlet is the cleaner test, and it delivers the same well-powered null reported above. Advertised salaries in the catchment do not move after hyperscale entry, at any ring, on a panel that averages several hundred salaried postings per cell. Under the standard mapping from wages to marginal product, no local productivity gain reaches incumbent workers. The growth outlet points the same way but identifies less cleanly. A within-site event study on the hiring activity of firms that existed prior to the hyperscale facility start shows no positive response at any horizon, with every event-study coefficient statistically indistinguishable from zero. The pooled and colocation samples carry a pre-existing downward trend in inner-versus-outer hiring that the design cannot fully remove, so we read the incumbent-hiring evidence as corroborating rather than identifying. Both outlets rule out the large positive productivity response that the agglomeration benchmark would predict.

5.5 Impact on realized county employment

The ring tests measure spillovers to firms near the facility, and they find none. A natural objection is that job postings are a thin measure. A data center's own jobs may be filled through corporate channels rather than posted to local boards, so a postings null could reflect measurement rather than the absence of jobs. We address this by turning to realized employment in standard county-level data, and by reading the effect through its industrial composition, which is what separates a direct effect from an agglomeration spillover. The composition is also what defends a county comparison against the siting selection that Section 3 documents. The confounds that make cross-county comparisons dangerous, such as an energy or land-development boom or a broad local upswing, would lift employment across many sectors at once. A data-center opening should instead raise employment in the facility's own sector and its network complement and nowhere else. Finding the effect confined to those two sectors, with every downstream and placebo sector flat, is the signature of the facility rather than of the place that attracted it.

We estimate the county event study of equation (3) on Quarterly Census of Employment and Wages employment, comparing counties that gain a hyperscale data center to counties that never do, for the data-center sector, its network complement, and the downstream industries that use compute. Table 2 reports the post-opening effect by industry and by facility type.

Table 2: County employment effect by industry and facility type

Industry (NAICS)	Hyperscale	Colocation	All data centers
Direct and network			
Data processing (518)	$+56^{***}$	$+37^{***}$	$+38^{***}$
Telecommunications (517)	$+43^{*}$	null	null
Downstream / compute-using			
Web and information services (519)	null	null	null
IT consulting (5415)	null	null	null
Professional, scientific, technical (541)	null	null	null

Notes: Post-opening log-point change in employment (approximately percent) from the county event study of equation (3), county and year fixed effects, never-treated counties as the comparison group, standard errors clustered on state (treated counties span 25 states). Pre-period leads for the direct sector (518) are jointly null ($p = 0.53$); leads in some downstream sectors are noisier, and the contested downstream sector (519) is corroborated with the staggered-robust estimator described in the text. "null" denotes a coefficient not distinguishable from zero at the ten-percent level. $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$.

The composition is unambiguous. Realized employment rises in the data-center sector itself and in telecommunications, the network complement that data centers carry with them. It does not rise in any downstream or compute-using industry, and this holds for hyperscale counties, colocation counties, and the full set of data-center counties. The positive employment a county-level design records is the facilities and their network, not a spillover to surrounding firms. The accompanying establishment counts confirm the reading. The rise in data-processing establishments is the data centers and colocation providers themselves, while establishment counts in the downstream industries are flat.

The web and information services sector (519) is the one downstream industry where a level-matching design can appear to show an effect. It does not survive a staggered-robust estimator. Under a Callaway–Sant'Anna design with not-yet-treated controls, the post-opening effect is small, statistically insignificant, and no larger than the sector's own pre-period leads, and its establishment count is flat throughout, so there is neither firm entry nor anticipatory co-location. The apparent signal is a pre-trend rather than an effect. This bears on our county-level companion paper [Bahar and Wright, 2026], which reported a downstream effect. That estimate does not survive the staggered-robust estimator used here, and the flat establishment count shows that the firm entry an agglomeration cluster would require never appears.

6 Incidence on land

A modest amount of value stays in the catchment, accruing to land rather than to labor. We find that residential rents rise near hyperscale facilities. The within-site rent cells are +6.6% at 0–5 km, +9.4% at 5–10 km, and +8.5% at 10–25 km, with the peak in the broader residential catchment rather than immediately at the campus. Two cautions pull the headline number down. The first is a synthetic control built for each facility from non-data-center tracts, which returns smaller per-facility magnitudes. The second is a placebo on comparison sites of similar density, which shows that random non-data-center locations also saw rent growth of about +8.4% at 5–10 km during the 2019–2023 work-from-home window. Part of the within-site coefficient therefore reflects a general densified-area pattern rather than a data-center effect. We read the data-center-specific rent residual as +2 to +5% at the 5–10 km peak, with the synthetic control as the headline and the within-site estimate as an upper bound. The built-versus-cancelled design, which cannot resolve fine spatial detail, is consistent with this reading. Median listing rents at built sites rise 1.9 percent relative to cancelled sites ($p = 0.59$) with a flat pre-trend ($p = 0.23$), and the cancelled sites' own rent path is flat. Colocation rent is null at every ring and in every robustness specification.

The rent rise arrives without the supply or demand signatures of a local boom. A place-level Census Building Permits Survey check finds no multifamily permitting response in the same ring, with 5-or-more-unit permits up only +3.4% ($p = 0.87$). On the demand side, a county-to-county migration event study on IRS Statistics of Income (SOI) data shows in-migration rising to +6.3% and adjusted-gross-income inflow to +9.0% by year five after a hyperscale opening, with a clean colocation null. But out-migration rises by a comparable amount over the same window, so the net change in residents is near zero. The gross flows describe churn and a shift in composition toward higher-income in-migrants, not population growth. With no new supply and no net new residents, the rent increase is best read as capitalization into the existing housing stock.

Foot traffic shows no robust change at built sites relative to cancelled ones. The inner-ring estimate from the built-versus-cancelled design is about −12% ($p = 0.06$) on the full panel, with a flat pre-trend ($p = 0.50$), but it falls to -3% ($p = 0.68$) once the panel is balanced to sites observed across the full event window, so we do not read it as a reliable displacement. The colocation foot-traffic decline, by contrast, runs on a smooth secular slope with no break at opening, so we do not read it as an effect either.

Two further pieces of evidence fit this reading. Local public finances strengthen after a hyperscale opening, with increased school revenues and higher capital outlay and public-works spending, so part of the measured rent rise plausibly capitalizes local fiscal capacity and infrastructure rather than labor-market demand. And commercial spending rises weakly at 0–5 km, by about +8.5% on the cleaned point-of-interest panel, though the within-site standard error leaves it imprecise ($p = 0.50$). It passes a random-donor placebo, but we do not treat it as a separately identified channel. The spending response is consistent in direction with the rent and migration evidence and equally consistent with capitalization rather than the formation of a new local consumer base.

Figure 2 reports the migration event study in full. The hyperscale in-migration rise survives a pairs cluster bootstrap (199 resamples of county fips), with in-migration of +6.3% at $p = 0.010$ and AGI inflow of +9.0% at $p < 0.005$ by year five, while the colocation in-migration cell stays null.

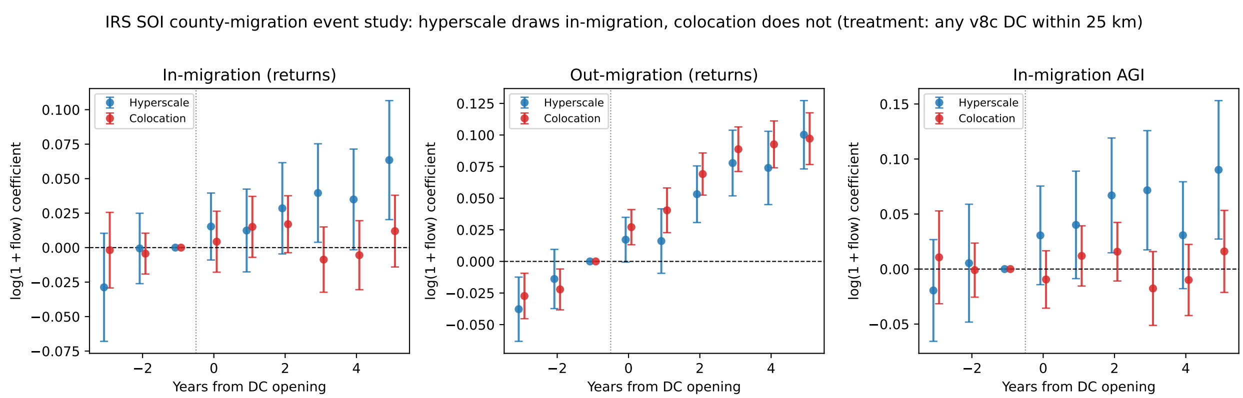

Figure 2: IRS SOI county-migration event study around DC openings. Hyperscale openings are followed by rising in-migration of returns and AGI, while colocation openings are null on in-migration. *Notes:* County-level event study of log return-inflow, log return-outflow, and log AGI-inflow around data-center openings (`cf_240`). Treatment: county is treated at year $t^\star$ if any high- or medium-confidence data center of the listed archetype sits within 25 km of the county centroid and opened in year $t^\star$. Hyperscale arm: 76 treated counties, 3,145 control counties; colocation arm: 191 treated counties, 3,030 control counties. Specification: $\log(1 + \text{flow})$ on event-time dummies d_e for $e \in \{-3, \ldots, +5\} \setminus \{-1\}$ with county and state $\times$ year fixed effects; CRV1 standard errors clustered on county fips. Reference period is $e = -1$. Inflow / outflow / AGI are the IRS "Total Migration-US" rows aggregating all US domestic flows into / out of each county-year. Markers offset ± 0.08 on the x-axis for visibility. 95% confidence intervals shown.

The fiscal evidence also reinforces the capitalization interpretation. Table 3 combines school-district and non-school local-government finance panels. Hyperscale openings are followed by larger local school revenues, higher school capital outlay per pupil, and higher county highway and public-works spending. Part of the measured rent rise is therefore consistent with households capitalizing local fiscal capacity and infrastructure improvements that arrive with hyperscale development.

Table 3: Public-goods capitalization checks

Outcome	Geography / sample	Hyperscale	Colocation
Local property-tax revenue	school district	+24.6%***	+8.9%***
Total local revenue	school district	+12.8%***	+3.8%*
Capital outlay per pupil	school district	+35.0%*	+4.3%
Property-tax revenue	place, balanced	+20.2%***	+10.6%***
Highway current + construction	county, balanced	+47.6%***	+10.9%
Broad public works current + construction	county, balanced	+18.7%**	+11.0%*
Public-building current + construction	county, balanced	+75.9%**	+27.2%

Notes: School-district estimates use NCES/Census F-33 finance files for fiscal years 2012–2023 (`cf_210`). Local-government estimates use Census Annual Survey/Census of Governments individual-unit files for fiscal years 2012–2023 (`cf_211`). All specifications use geography fixed effects and state-by-year fixed effects, with standard errors clustered by geography. The local-government rows report the balanced 2012–2023 panel. Public-works outcomes use the consistently available current plus construction margin (Census item families E** + F**); other-capital-outlay G** codes are excluded because they do not appear consistently in the 2022–2023 public-use files. Stars: $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$.

The incidence is therefore narrow and lands on owners. The surplus that stays local shows up in higher land and housing values and in stronger municipal balance sheets. It is captured by landowners and local governments, alongside the operator's own capitalized footprint in the nighttime-lights signal. Local workers and local firms, the parties an agglomeration spillover would benefit, are left roughly where they started.

7 Incidence on electricity rates

The most common objection in data-center siting debates is not about jobs or land but about electricity bills. Opponents claim that a hyperscale facility's enormous load will raise rates for the households that share its grid. The claim has national versions as well. Wholesale-market modeling finds that existing data centers have raised average wholesale prices by 3 to 5 percent, with larger effects in data-center corridors [Taylor et al., 2026]. Whether the households and businesses in a host utility's territory pay more, however, is a distinct question, and the existing evidence on it is associational. Data centers select territories with cheap power, so a comparison of rates between utilities with and without data centers conflates the siting choice with its consequences. Our cancelled-facility cohort provides a useful counterfactual, namely utilities whose territories were selected for a hyperscale project on the same fundamentals but where the facility was withdrawn or lost the site contest.

We measure retail rates from the Energy Information Administration's Form 861, computing each utility's average residential price as revenue over megawatt-hours sold for utilities in each state and year from 2012 to 2024. A utility is treated in the year the first facility of a given archetype in our registry opens in a county it serves, and we run the hyperscale and colocation archetypes as separate arms.

To keep the arms clean we exclude utilities that host both archetypes from treatment, and control utilities host no facility of any registry type, including the smaller enterprise and crypto sites outside our two archetypes. The hyperscale arm compares 54 treated utility-states with first openings between 2015 and 2021 to roughly 2,000 never-treated utilities and, separately, to the 54 near-miss utilities. The colocation arm has 60 treated utility-states. Utilities first treated in 2014, when our satellite-based dating is left-censored, and after 2021, with too little post-period, are excluded from both groups. Because treatment is assigned at the county level, utilities serving a different corner of a treated county are coded as treated, which biases the estimates toward zero, so the design is conservative.

Figure 3 shows the result. In the hyperscale arm pre-trends are flat, with pre-period coefficients averaging 0.0% against never-treated controls and −0.3% against the near-miss controls, so the siting concern that motivates the near-miss comparison does not in fact manifest as differential price trends. After a hyperscale opening, residential rates do not rise. They fall. Against never-treated controls with state-by-year fixed effects the decline builds steadily, to −2.3% two years after opening ($p = 0.03$), −3.1% at year three ($p = 0.02$), and −4.0% at year five ($p = 0.03$).

The near-miss comparison, with only 54 control utilities, is noisier but ends in the same place, at −3.2% by year five ($p = 0.11$). Its control pool is too small to support state-by-year fixed effects, so that arm uses year fixed effects, and its value is the flat pre-trends rather than the tightest standard errors. A broader specification that keeps all 90 hyperscale-hosting utilities regardless of colocation presence reaches −4.1% at year five against near-miss controls ($p = 0.03$). The results survive restricting to a balanced panel of utilities observed in at least twelve of the thirteen years, which addresses a change in EIA's coverage of small utilities in 2020. On the balanced panel the year-five effect is −3.1% ($p = 0.08$) against never-treated controls and −3.0% against the near-miss controls, on 33 treated utilities. Commercial rates show the same null-to-negative pattern against never-treated controls (−1.2% at year five, imprecise), and industrial rates, though noisier, show no increase against the near-miss controls (−1.9% at year five, $p = 0.67$).

The colocation arm is the specificity check, and it behaves exactly as the load-based interpretation requires. Colocation openings are followed by no rate decline at any horizon. The point estimates are mildly positive and never significant (+2.1% at year five, $p = 0.32$). A colocation facility adds racks inside an existing building envelope and draws a fraction of a hyperscale campus's load, so it adds little volume over which to spread the distribution network's fixed costs. The same hyperscale–colocation asymmetry that runs through the nighttime-lights, rent, and migration results therefore reappears in electricity rates, and it ties the rate decline to the one thing hyperscale facilities bring that colocations do not, which is load.

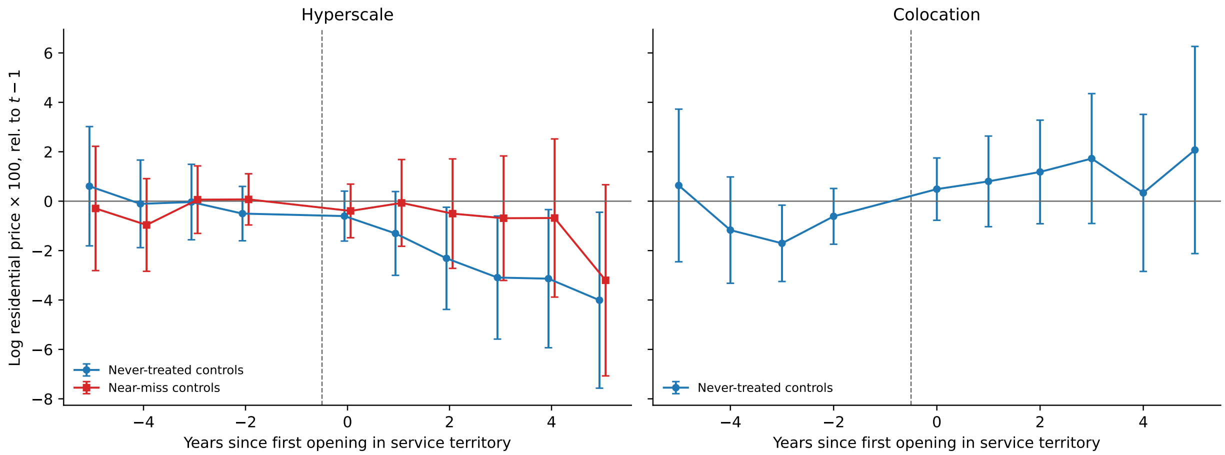

Figure 3: Residential electricity prices around the first data-center opening in the serving utility's territory, by archetype. Rates decline by roughly 3 to 4% over five years after a hyperscale opening, while colocation openings show no effect.

Notes: Event studies of log average residential price (revenue over MWh, EIA Form 861 bundled-service utilities, 2012–2024) around the first opening of each data-center archetype in a county the utility serves. Left panel, hyperscale arm: 54 treated utility-states with first openings 2015–2021 and no colocation facility; blue series uses 2,040 never-treated controls (no facility of any registry type, including non-archetype enterprise and crypto sites) with utility and state×year fixed effects, red series uses 54 near-miss controls (utilities serving counties where an announced hyperscale project was cancelled or withdrawn, same no-facility restriction) with utility and year fixed effects. Right panel, colocation arm: 60 treated utility-states with first colocation (tier-1 or broader) openings 2015–2021 and no hyperscale facility, against the same never-treated controls with utility and state×year fixed effects. Utilities first treated in 2014 (left-censored dating) or after 2021 are excluded from treatment and control; utilities with fewer than 1,000 residential customers excluded. CRV1 standard errors clustered by utility-state; 95% confidence intervals; reference period $t-1$. Endpoint bins are clipped at ± 5 and unbalanced across cohorts.

The two findings, falling local rates here and rising wholesale prices in Taylor et al. [2026], are not in tension. They describe different channels with different geographic incidence. A utility-level comparison nets out price movements common to a wholesale market region, because treated and control utilities buy power from the same market. What remains is the local distribution channel, and there the arithmetic favors incumbent ratepayers. A hyperscale facility adds enormous volume over which the fixed costs of the distribution network are spread, and it typically does so under a special contract or dedicated large-power tariff. Because that load is billed outside the residential class, the facility's own consumption does not enter the residential revenue-per-megawatt-hour we measure, so the decline is not a mechanical artifact of adding the data center's purchases to the denominator. The finding also bears on the cost-shifting concern now before several state utility commissions, namely that discounted data-center tariffs shift network costs onto household bills. In the host territories we study, residential rates fall rather than rise through 2024 after a hyperscale opening, so whatever reallocation occurs inside these utilities' rate structures does not appear as higher household rates over this horizon. The wholesale channel, by contrast, is borne by everyone in the market region, in territories with and without data centers alike. The local political economy of data-center siting therefore has the incidence roughly backwards. The rate consequences a community can affect by refusing a facility are, if anything, slightly favorable to hosting it, while the consequences it cannot affect arrive through the regional market regardless of where the facility lands. We note one limit on scope. Our registry's dated openings thin out after 2022, so these estimates describe the cloud-era build-out through 2021 cohorts followed to 2024, not the post-2022 AI surge, whose wholesale-side effects are the subject of Taylor et al. [2026].

8 Interpretation

Table 4 summarizes all of the outcomes across rings for both hyperscale and colocation, derived from the within-site ring difference-in-differences (DiD).

Table 4: Outcome matrix: hyperscale versus colocation

	Hyperscale				Colocation			
Outcome	0–1	0–5	5–10	10–25	0–1	0–5	5–10	10–25
Capital								
NTL (log radiance)	+40%***	+6%***	null	null	+5%***	+1%	−1%	−1%***
Displacement								
Foot traffic (visits)	—	+1%	+2%	−1%	—	−9%***	−8%**	−4%
Agglomeration								
Commercial spending	—	+8.5%	+10.4%	+9.9%	—	null	+4.8%	−3.0%
Median rent	—	+6.6%**	+9.4%***	+8.5%***	—	null	null	null
Workplace employment†	—	null	null	null	—	+2%***	+2%**	null
High-earnings share	—	+0.20pp*	—	—	—	−0.47pp**	—	—
Agglomeration-channel tests (this paper)								
Supplier-sector postings	—	null	null	null	—	null	null	null
Advertised wage	—	null	null	null	—	null	null	null
New-firm births	—	null	—	—	—	—	—	—

Notes: Within-site ring difference-in-differences (equation 2) on the harmonized sample of 1,113 data centers (339 hyperscale), each outcome over a common 24-month post-opening window against the 25–50 km outer ring. The 0–1 km column is the native inner ring; the 0–5 km column area-weights the 0–1, 1–2, and 2–5 km rings, so the capital footprint that is +40% at 0–1 km (+5% for colocation) averages to +6% at 0–5 km. Only the capital footprint is identified at the 0–1 km ring, and "—" marks cells not estimated at a given ring. NTL uses facility-by-ring and calendar-quarter fixed effects, a two-period near-versus-outer difference-in-differences; foot traffic, spending, and rent use state-by-month fixed effects, which absorb the 2020–2023 work-from-home shock; employment uses year fixed effects. Spending and rent are detrended; the data-center-specific rent rise is bracketed to 2 to 5 percent by a synthetic control (Section 6). †Workplace employment fails the ring design's pre-trend test (Section 3); the cell is shown for completeness, and the paper's conclusion for it rests on the county employment design, not on this estimate. "null" denotes a coefficient not distinguishable from zero. $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$.

The matrix reports every outcome at the same coarse rings, including the null cells, so the ring-decay pattern is visible for each. The only outcome present at the inner ring is the capital footprint. The high-earnings share row measures composition rather than scale. It is the share of workplace jobs in the LODES high-earnings band, with monthly earnings above $3,333, and its small positive hyperscale cell and negative colocation cell indicate mild shifts in the mix of jobs rather than job creation. The three direct agglomeration-channel tests in the bottom panel, namely supplier-sector postings, advertised wages, and new-firm births, are null at every ring. The job-postings and employment cells are discussed in Section 5, and the rent, spending, and foot-traffic cells in Section 6.

Two caveats temper the matrix as a clean archetype comparison. Hyperscale and colocation facilities sit in structurally different places. The median 25 km catchment is 510 to 780 thousand people for hyperscale and 1.2 to 1.5 million for colocation, and hyperscale campuses sit a median 46 to 66 km from a top-50 metro centroid against 23 km for colocation. Some of what reads as an archetype difference may reflect the kind of place each prefers rather than the kind of facility it is. And most outcome windows overlap the 2020 to 2023 work-from-home period. The state-by-month fixed effects absorb it, but part of the colocation foot-traffic decline is a secular trend rather than a facility effect (Section 6).

The pattern is simple to organize. A large plant opening reaches the local economy through three agglomeration channels, the shared labor pool, supplier linkages, and knowledge spillovers among nearby firms. Each channel requires either local labor or local input-output linkages, and a data center supplies neither. It employs dozens of workers, not thousands, so it deepens no labor pool. It buys electricity and fiber through national markets, so it localizes no suppliers. It ships compute over the cloud to distant customers, so it seeds no knowledge cluster. The channels are absent by construction, which is why a footprint this large does not propagate.

What remains is the capital itself and a narrow incidence on land. The footprint is the facility's own physical capital, large and confined to the inner kilometer. Its only local transmission runs through housing, a modest rent increase in the catchment that capitalizes into the existing stock. The transmission is small because the new residents are few, and it shows up as a price rather than a quantity because the housing stock is fixed within the horizon we observe. The relevant constraint is the size of the existing stock relative to the shock, not the long-run supply elasticity, which has no time to operate. This is why the percentage rent response is larger in small exurban housing markets, where hyperscale campuses tend to locate, than in dense metros.

The contrast between the two facility types follows the same logic. Hyperscale facilities are large greenfield builds, so they show the capital footprint and the rent response. Colocation facilities retrofit existing space and add capacity rather than construction, so they show neither. Because the agglomeration channels are absent by the structure of the technology rather than by happenstance, the conclusion is unlikely to reverse as the sector grows.

9 Conclusion

A hyperscale data center is as large a capital investment as the manufacturing plants that anchor the agglomeration literature, but our evidence shows it does not behave like one. The capital footprint is large and local, and it stops there. Relative to announced-then-cancelled sites, brightness at built sites rises sharply at groundbreaking and fades to nothing within five kilometers, the within-site ring design reproduces that jump on a flat pre-trend, and the parcel's vegetation clears in the same year the lights appear. None of the three channels through which a large plant raises local productivity shows a detectable response. The local labor pool does not deepen, suppliers do not localize, the downstream compute industry does not cluster, and new-firm registration shows no increase. In the realized county employment data, the gains are confined to the data centers and their direct network. Colocation facilities, which add capital without new construction, are null on every margin.

The only local outcome that moves with the capital is land. Residential rents rise by a few percent at built sites relative to cancelled ones, whose own rent path stays flat, the rise arrives without new multifamily permitting and without net in-migration, and gross in-migration is offset by comparable out-migration, so the increase capitalizes into the existing housing stock. Foot traffic shows no robust change relative to cancelled sites, and commercial spending does not move. The local incidence is narrow. It falls on landowners and on municipal balance sheets, not on local workers or firms.

Three implications follow for place-based incentives. A jurisdiction that recruits a hyperscale facility should expect a large physical footprint and a modest capitalization into land values, but not the labor-market or supplier spillovers that justify subsidies aimed at conventional plants. The roughly twenty billion dollars committed to attracting these facilities is buying capital, not agglomeration. The incidence falls on owners of land and on local public finances rather than on workers, which matters for who gains and who pays when a subsidy is financed locally. And because the agglomeration channels are absent by the structure of the technology rather than by chance, the result is unlikely to reverse as the sector grows. The local-multiplier logic that animates industrial policy does not extend to labor-light digital capital. Our estimates describe the cloud-era build-out through 2021 cohorts. The post-2022 artificial-intelligence campuses are larger still. Their scale should strengthen the load-driven rate result, and it does not change the structural absence of labor linkages.

A The instrument search

A natural way to address the siting-selection problem is an instrument for data center location. We searched for one and did not find a clean one. This appendix documents the search because the failure is informative. It is the reason we identify the effect from the within-site and county designs rather than from an instrument, and it characterizes how hyperscale siting actually works.

Predetermined siting instruments. We tested each plausibly exogenous, predetermined predictor of data-center siting, interacted where needed with the national capacity wave, and screened each against a manufacturing placebo (a valid instrument for data-center activity should not predict manufacturing employment). Table 5 summarizes the outcomes. Transmission-substation density predicts siting strongly but fails the placebo, because it is a proxy for industrial land. The 1980 urban college share, which works in the all-data-center county literature, carries the wrong sign for hyperscale, because hyperscale campuses go to low-college exurban land rather than to educated metros. Long-haul fiber proximity predicts siting only across metropolitan areas, not within a county, so it cannot identify a sub-county effect. Climate, measured by wet-bulb temperature and free-cooling days, does not predict siting at all, because the largest US hyperscale hub is hot and humid and operators engineer around climate to chase cheap power. Generation-capacity timing has no within-county predictive power. The pattern is systematic rather than bad luck. The features that predict hyperscale siting (power, industrial land, fiber) are the same features that are correlated with local industrial character, and the features that are clean (climate, historical human capital) do not predict where these facilities go. Hyperscale siting is cost-driven, and there is no clean predetermined instrument for it at sub-county scale.

Table 5: Predetermined siting instruments and why each fails

Instrument	First stage	Why it fails
345 kV substation density	strong	fails manufacturing placebo (industrial proxy)
1980 urban college share	moderate	wrong sign (hyperscale → low-college exurbs)
InterTubes fiber proximity	metro-scale	null within county
Climate (wet-bulb, free-cooling)	null	siting is cost-driven; the main hub is hot-humid
Generation-capacity timing	null	no within-county power

Notes: Each instrument is the predetermined county feature interacted with the national hyperscale capacity wave. "Fails placebo" means the instrument predicts manufacturing employment, violating the exclusion restriction.

A shift-share instrument is also weak. A shift-share design of the form used in concurrent county-level work [Alvarez et al., 2026], interacting the predetermined fiber and college exposures with national data-center demand growth, fares no better here, and the obstacle is the endogenous variable. Our public measures of data-center scale, namely capacity, square footage, and facility count, are lumpy and supply-driven, and a demand-side shift predicts a lumpy supply-side object poorly. The first stage with square footage as the endogenous reaches an F of only twenty to thirty, too weak to use.

Because the first stage is weak, we do not use its estimates as evidence. We note only, descriptively, that nothing in the instrumented results contradicts the paper's composition finding. The two-stage least-squares point estimates load on the direct data-center sector and are null or negative in the downstream industries. The instrument also loads on population-serving sectors such as health, the general area-development confound that the within-site and cancelled designs avoid, which is a further reason we treat the exercise as documentation of the search rather than as a design.

B The geography of compute

A common rationale behind data center incentives is that the facility will anchor a local digital economy. If the cloud servers are here, the reasoning goes, the firms that use the cloud will follow. This section examines that premise directly by mapping where the heavy users of compute actually sit relative to data centers, and by asking whether their location reflects the data center or something the data center does not provide. We use the Veridion firmographic database [Veridion, 2026], which geolocates 6.5 million US establishments and classifies each by industry, employment, and founding year. We define compute users by primary industry. Software publishing, computing infrastructure and hosting, computer systems design, and web and streaming services form the compute-native core, with finance, scientific R&D, and semiconductors as a broader set.

B.1 Compute users cluster near data centers

Compute users are closer to data centers than the average firm. The median compute-native firm sits 11 kilometers from the nearest data center, against 20 kilometers for non-compute firms, and 69 percent of compute users fall within 25 kilometers of a data center, against 55 percent of other firms (Table 6). The gradient runs as one would expect from the technology. Cloud and hosting firms are nearest at a 6 kilometer median, software firms next at 9 kilometers, while telecommunications and streaming, whose footprints are more dispersed, sit farthest out.

This proximity is not an artifact of headquarters location. Measuring each firm's operating footprint by the geography of its job postings rather than its headquarters returns the same pattern at a slightly larger distance, because a firm's branch offices spread beyond its headquarters metro. Nor is it merely that both compute users and data centers crowd into a handful of tech metros. Removing the 45 largest metropolitan areas, compute users remain closer to the surviving data centers than non-compute firms are (a 75 versus 87 kilometer median, with software firms among the closest), so the association holds in second-tier and rural America as well.

Table 6: Distance from firms to the nearest data center

	Firms	Median km	$\leq$ 25 km	$\leq$ 50 km
Panel A: Headquarters location (Veridion)				
Non-compute firms	5,692,641	20.5	55%	70%
Compute users (core)	309,760	11.0	69%	81%
Cloud and hosting	26,188	6.1	75%	84%
Software	101,757	8.9	72%	84%
Computer systems design	154,934	11.7	68%	81%
Finance	193,540	15.9	62%	75%
Telecommunications	10,125	22.2	53%	66%
Panel B: Establishment footprint (job postings)				
Compute users (core)		15.1	63%	76%
Panel C: Outside the 45 largest metros				
Non-compute firms	2,603,195	86.8	20%	33%
Compute users (core)	103,356	75.3	30%	41%

Notes: Distance from each firm to the nearest hyperscale-clean data center, by haversine. Panel A places each firm at its Veridion headquarters coordinate. Panel B replaces the headquarters with the firm's job-posting locations (WageScape), weighted by the number of distinct hiring firms at each location, so it reflects the operating footprint rather than the headquarters. Panel C excludes all firms and data centers within 50 km of the 45 largest metropolitan and technology-hub centers and measures distance to the nearest remaining data center. Compute-native "core" sectors are software publishing, computing infrastructure and hosting, computer systems design, and web and streaming services.

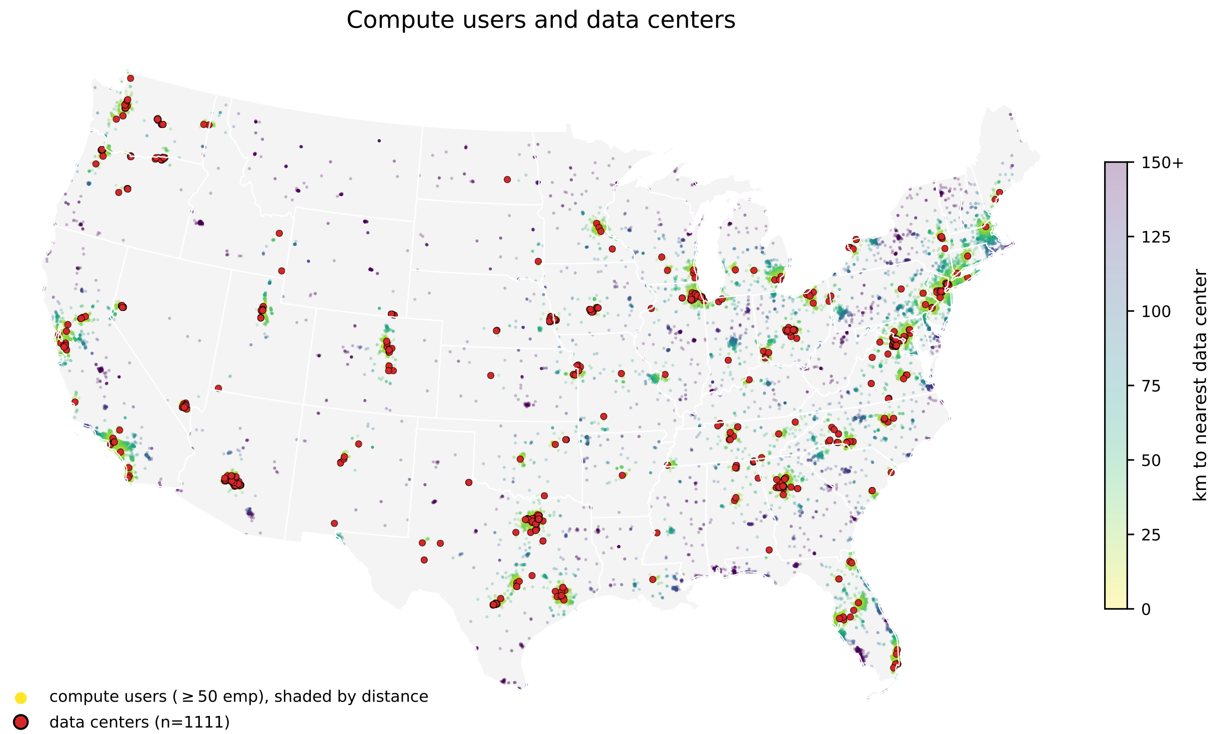

Figure 4: Compute-user firms and data centers across the continental United States. Points are compute-native firms with at least 50 employees, shaded by distance to the nearest data center, with data centers marked in red. Both concentrate in the same coastal and metropolitan corridors, while interior firms shade dark, far from any facility.

The two populations trace the same map (Figure 4), crowding into the coastal and metropolitan corridors and thinning together across the interior. The pull is strongest for the largest firms. When sorted by employment, the median distance to a data center falls monotonically from 13 kilometers for the smallest firms to 4 kilometers for those with more than ten thousand employees, and the dispersion collapses with it. The biggest compute users cluster tightly around data centers, and almost none are far from one. Only about 5 percent of compute firms with at least a thousand employees sit more than 50 kilometers from any data center. The handful that do are software companies that scaled in place in second-tier cities, such as Esri in Redlands, Jack Henry in rural Missouri, and Ansys outside Pittsburgh. Distance to compute is a constraint these firms evidently did not face, because the cloud let them consume capacity hundreds of kilometers away.

B.2 Two kinds of data center hub

The co-location is a metropolitan phenomenon, and seeing this matters for what follows. Grouping data centers into hubs reveals a sharp split. Metropolitan hubs are embedded in dense clusters of sector-matched compute users. The 239 data centers around Ashburn, Virginia sit among thousands of government IT contractors, with DXC, Leidos, Booz Allen, SAIC, and CACI all within 25 kilometers. The Bay Area hubs are ringed by Google, Broadcom, Synopsys, and ServiceNow. By contrast, the large exurban hyperscale clusters are isolated. The 50 data centers around Hillsboro, Oregon and the cluster near Quincy, Washington each have on the order of four compute firms within 25 kilometers. The same physical infrastructure produces a thick compute neighborhood in a tech metro and an empty one in the exurbs.

B.3 Why they co-locate, and why it is not the data center

Part of the explanation is shared infrastructure. Data centers and compute users both require access to the high-voltage grid, and both sit closer to it than ordinary firms. The median data center is 3.4 kilometers from a 230-kilovolt substation, the median compute user 5.0 kilometers, and the median other firm 6.4 kilometers. High-voltage substation density independently predicts where compute users locate. But the grid is only part of the story. The proximity between compute users and data centers survives controlling for substation access, so a common pull toward the power backbone, together with the fiber and talent that concentrate in metros, leaves the two in the same places, consistent with shared infrastructure rather than one drawing the other.

Three pieces of evidence indicate that the data center is not the cause. First, the compute users near data centers predate them. Among compute firms within 25 kilometers of a data center, at least 69 percent were founded before the nearest data center opened. Because the firm registry over-represents recently founded firms, the true share is higher. The local compute base existed first. Ashburn illustrates the direction of causation. The government technology ecosystem of Northern Virginia and its early internet exchange drew the data centers to the region, not the reverse. Second, the timing evidence in Section 3.5 shows that data center openings do not raise job postings in the surrounding rings, for compute-using industries or any other, once a smooth pre-existing trend is removed. The opening is not a turning point for local hiring. Third, and most directly, exurban data centers sit in compute deserts and stay there. The median metropolitan data center has roughly 2,700 compute firms within 25 kilometers, while the median exurban data center has 41 (Table 7). The facility does not move that number.

Table 7: Compute users near metropolitan versus exurban data centers

	Data centers	Median compute firms ≤ 25 km	Share with < 5 nearby
Metropolitan data centers	824	2,702	0%
Exurban data centers	287	41	11%

Notes: Each data center is classified as metropolitan if it lies within 50 km of one of the 45 largest metropolitan or technology-hub centers, and exurban otherwise. The middle column reports the median count of compute-native firms (Veridion) within 25 km of the data center; the last column the share of data centers with fewer than five compute firms within 25 km. Separately, among compute firms within 25 km of any data center, 69 percent were founded before the nearest data center opened.

B.4 Implication for place-based policy

The policy reading is direct. A jurisdiction weighing a data center, especially an exurban one, should not expect compute users to arrive in its wake. Firms that consume cloud capacity locate near transmission, fiber, and skilled labor. These assets concentrate in metropolitan areas, and a single data center does not create them. The compute users already near data centers were mostly there first, and the data centers that open in compute-sparse places remain in compute-sparse places. What the facility delivers locally is the capital footprint and the rent pressure documented above, not a cluster of digital firms. The hope that the servers will bring the software is, on this evidence, misplaced.

C Supplementary results

Table 8 reports every outcome under every design in which it can be estimated, so the routing rule of Section 3 can be audited cell by cell rather than taken on faith.

Table 8: Every outcome under every design: estimates and routing

Outcome (hyperscale, inner ring)	Built vs. cancelled	Within-site rings	County event study
Capital (nighttime lights)	**+68% ($p < 0.01$)**	**+38% ($p < 0.01$)**	—[a]
Residential rent	+1.9% ($p = 0.59$)	**+2 to +5%**[b]	—[a]
Advertised salary	+1.7% ($p = 0.92$)	−1.3% ($p = 0.65$)	—[a]
Job postings	**+9.1% ($p = 0.46$)**[c]	excluded[d]	—[a]
Foot traffic	−11.5% ($p = 0.06$)[e]	excluded[d]	—[a]
Consumer spending	**+7.2% ($p = 0.38$)**	+8.5% ($p = 0.50$)[f]	—[a]
Firm entry (SoS registrations)	—[g]	−0.7% ($p = 0.46$)[h]	+1.4% ($p = 0.49$)[i]
Workplace employment	—[g]	excluded[d]	**518: +56 log pts***; downstream: null**

Notes: Post-event estimates for the hyperscale arm at the innermost ring each design supports (0–1 km for nighttime lights, 0–5 kr otherwise). Bold marks the design that carries the paper's interpretation for that outcome. Built-versus-cancelled capital estimate i the isolated-sample groundbreaking-year jump; magnitudes at longer horizons rest on few cancelled controls (Section 4).

[a] Not estimated at the county level; no county analog of the ring outcome.

[b] Synthetic-control headline; the raw within-site cells (+6.6 to +9.4%) are an upper bound (Section 6).

[c] Robust across a precision family: Poisson counts +11.3% ($p = 0.45$), long difference −6.6% ($p = 0.67$), construction-dated +3.0% ($p = 0.82$). Minimum detectable increase at conventional power is roughly forty percent.

[d] Fails the flat-leads gate: a pre-existing inner-versus-outer trend the ring design cannot separate from an effect (Section 3).

[e] Falls to −3.0% ($p = 0.68$) on the balanced panel, so not read as a reliable displacement.

[f] Reported with caveats in Section 6; passes a random-donor placebo.

[g] No cancelled-arm panel exists for this outcome (business registrations and LODES cannot be reconstructed around never-built sites a ring scale).

[h] Hyperscale-only sample; pooled all-archetype estimate +0.5% ($p = 0.42$), with offsetting state heterogeneity reported in the appendix

[i] Census Business Formation Statistics, national county design.

C.1 Phase decomposition: construction versus operations

The capital-footprint estimates in Section 4 treat the nighttime-lights footprint as a single object dated to construction start. We can split it further, into the part that accrues during the build and the part that arrives with operations. Roughly one quarter of the hyperscale footprint accrues during the construction phase, before operations begin, and the remaining three quarters arrives when operations come online. This is the central reading of Table 9.

We anchor the construction-start date on three independent sources. The first is the sustained drop in the Sentinel-2 normalized difference vegetation index (NDVI) that marks land clearing on the facility footprint (`cf_097`). The second is the earliest FAA Obstruction Evaluation Form 7460-1 filing within 500 meters of the facility, restricted to construction-relevant structure types (cranes, transmission-line towers, buildings, water systems) and to filings within a plausible window around opening (`cf_310`). The third uses the same FAA filter but at a 1 kilometer radius for a larger sample at the cost of admitting some neighbor-project contamination. The three sources draw on different sensing modalities (optical satellite, regulatory filing) and different spatial radii, so convergent estimates are informative about the underlying object rather than about any one timestamp's measurement properties.

We replace the single post-opening indicator in the within-site DiD with two phase indicators. `post_construction` equals one for facility i in year-quarter t when t is on or after the construction-start date and before the operations date, and `post_operations` equals one when t is on or after the operations date. We interact each with the inner-ring indicator and estimate the two interactions jointly on the standard near-versus-outer panel (0 to 5 km inner, 25 to 50 km outer). The construction-phase ATT is the coefficient on `near` $\times$ `post_construction`, the operations-incremental ATT is the coefficient on `near` $\times$ `post_operations`, and the operations-total ATT is their sum, with the joint standard error from the variance-covariance matrix.

The three samples produce similar point estimates but differ in precision. On the NDVI-dated sample (`cf_300`, $n = 43$ hyperscale), the construction-phase ATT is +8.0% ($p = 0.015$) and the operations-total ATT is +29.5% ($p < 0.001$). On the FAA-dated 500 meter sample (`cf_300b`, $n = 64$), the construction-phase point estimate is similar at +8.7% but not statistically distinguishable from zero ($p = 0.14$), and the operations-total ATT is +33.3% ($p = 0.03$). On the FAA-dated 1 kilometer sample ($n = 119$), the construction-phase ATT is +6.2% ($p = 0.07$) and the operations-total ATT is +28.4% ($p = 0.001$). The point estimates lie in a tight band, with the construction phase between 6 and 9 percent and the operations total between 28 and 33 percent. The implied construction share of the total is 0.22 to 0.27 across all three samples. The colocation rows in Panel B are statistically indistinguishable from zero on both phases in every sample. The decomposition therefore inherits the matrix's archetype contrast, in that hyperscale generates a sizable pre-operations footprint and colocation does not.

Two caveats temper the construction-phase point estimate. First, the FAA `receivedDate` is a permit filing date that precedes ground-breaking by roughly 3 to 12 months, so the FAA construction window includes some pre-construction calendar months in which no nighttime-lights change is expected. This biases the FAA construction-phase estimate downward relative to the bulldozers-to-operations true ATT and partially accounts for the larger standard errors in the FAA samples relative to NDVI. Second, identification of the phase decomposition leans on crossfacility variation in the construction-to-operations lead time. The interquartile range of lead times remains wide in all three samples (11 to 50 months at the FAA 1 kilometer anchor), which supports identification, but tight clusters of lead times around a single duration would not.

The decomposition has a substantive interpretation. The construction-phase NTL signal is consistent with on-site activity that begins before commercial operations, including perimeter security lighting, substation and switchgear installation, cooling-tower assembly, fiber-conduit work, and the construction-crew lighting that accompanies a multi-year hyperscale build. The operations-phase increment captures the remaining lift after commercial illumination, server-floor load, and continuous twenty-four-hour cooling come online. Roughly three quarters of the total within-site footprint is operations, and roughly one quarter is construction. The construction signal is temporary and ends when the facility opens, while the operations signal persists for as long as the facility runs.

Table 9: Phase decomposition of the hyperscale within-site nighttime-lights footprint

Timing source	n_{DC}	Construction-phase ATT (%)	Operations total ATT (%)	Construction share of total
Panel A: Hyperscale				
Sentinel-2 NDVI clearing	43	$+8.0^{**}$	$+29.5^{***}$	0.27
		(3.4)	(11.0)	
FAA OE permit, 500 m	64	+8.7	$+33.3^{**}$	0.26
		(6.0)	(15.4)	
FAA OE permit, 1 km	119	$+6.2^{*}$	$+28.4^{***}$	0.22
		(3.4)	(8.4)	
Panel B: Colocation				
Sentinel-2 NDVI clearing	21	−1.0	−1.4	—
		(2.3)	(4.3)	
FAA OE permit, 500 m	45	−5.3	−9.2	—
		(3.7)	(7.2)	
FAA OE permit, 1 km	120	−2.6	−3.6	—
		(2.1)	(3.9)	

Notes: Each row is a separate within-site ring difference-in-differences with two phase indicators replacing the single post-opening indicator: `post_construction` equals one between the construction-start date and operations, and `post_operations` equals one after operations begin. Coefficients are reported in percent ($100 \times (e^{\beta} - 1)$). Operations-total ATT is the sum of the construction-phase and incremental-operations coefficients, with standard errors computed from the joint variance-covariance matrix. Construction-start dates come from three independent sources: Sentinel-2 NDVI sustained vegetation drop (cf_097), and the earliest project-tied FAA OE Form 7460-1 filing within 500 m or 1 km of the facility (cf_310). Standard errors are clustered on dc_id. The 500 m FAA sample excludes ambiguous matches; the 1 km sample is the larger but noisier comparator. The downward bias on the construction-phase point estimate in the FAA samples reflects that the permit-filing date precedes ground-breaking by roughly 3 to 12 months. Construction share = construction-phase ATT divided by operations-total ATT. *, **, *** indicate $p < 0.10$, 0.05, 0.01.

C.2 New business formation: a contextual heterogeneity finding

A natural follow-up question is whether the commercial activity around hyperscale openings is driven by *new* firms entering the catchment or by *existing* firms expanding. Neither the SafeGraph Spend panel (transactions at existing POIs) nor the LinkUp postings panel speaks directly to firm entry. To address this, we assembled state-level business-registration panels for Virginia, Texas, and Iowa. These are the three DC-hosting states where bulk Secretary of State (SoS) data are freely available. We then ran the same within-site ring DiD on annual new-LLC formation counts at the fine and coarse rings (`cf_220`–`cf_228`). For Virginia and Texas we geocoded ∼1.4M and ∼890k unique addresses via the Census batch geocoder. Iowa publishes already-geocoded home-office coordinates. We also pulled the Census Business Formation Statistics (BFS) county annual file for a national county-level robustness check.

The result is a sharp contextual divergence:

Table 10: New business formation around hyperscale openings, by state and ring

State	Setting	n HS DCs	1–2 km		2–5 km	
			detrend	p	detrend	p
Virginia	dense exurban (NoVA)†	108	−3.6%	0.023	−3.8%	0.016
Texas	mixed Austin/Dallas	14	+5.2%	0.58	−9.2%	0.31
Iowa	rural Midwest	23	+11.8%	0.043	+10.6%	0.15
National county-level BFS:						
HS host counties	—	38 vs 2,520 controls	+1.4% detrended ($p = 0.49$)			

Notes: VA from Virginia SoS LLC filings 2014–2022, geocoded via Census batch (`cf_221`: 396,937 LLCs, 92% match rate within the 50 km catchment). TX from Texas franchise tax taxpayer addresses 2014–2026, geocoded similarly (`cf_223`: 1.54M filings, 90% match rate). IA from data.iowa.gov active business entity list 2014–2026 with pre-geocoded `POINT(lon lat)` home-office coordinates (`cf_227`: 211k active entities post-2014; the IA dataset is active-only, which is the binding measurement caveat for this row). National county-level row from Census BFS county annual file 2005–2024 (`cf_226`), where "host" is a county with any hyperscale DC and "control" is a non-DC county ¿50 km from any DC. All cells from within-site ring DiD (same as the matrix outcomes in Table 4). † NoVA = Northern Virginia. $^{*},^{**},^{***}$: $p < 0.10, 0.05, 0.01$.

This is a flow measure of new formations, not exits. The dependent variable is the count of newly registered LLCs and corporations in each ring-month cell. We do *not* observe closures or dissolutions of existing businesses. The Virginia −3.6 to −3.8% in the 1–5 km ring is therefore a reduction in the rate at which would-be operators choose to form a new entity in the immediate commercial neighborhood. It is not evidence that existing businesses are closing. The Iowa +11.8% is, symmetrically, an increase in the formation flow rather than a stock-level expansion of incumbents.

Three plausible mechanisms in the exurban case are consistent with reduced entry, namely commercial-rent capitalization, parcel-availability constraints from the DC footprint and supplier buildout, and the demand shadow of a small on-site DC workforce. Our identification does not separate them. The result fits the incidence evidence in Section 6. The same hyperscale openings that draw gross in-migration (Figure 2) and lift residential rents also suppress the small-entity entry margin in the immediate commercial neighborhood, with the negative effect concentrated where the rent and migration pressures are largest.

The three-state pattern in Table 10 is suggestive of a gradient in the local rent response, but should be read with two qualifications. First, with only three observations, the cross-state ordering is suggestive rather than estimated. A different set of states could overturn the sign. Second, the Iowa cell is the most fragile of the three. The Iowa Secretary of State dataset is an active-only registry, so dissolved entities are absent. The +11.8% at 1–2 km therefore partly reflects survivorship rather than entry alone.

With those qualifications, the three states sort along a plausible rent-pressure gradient. Virginia, where the within-site hyperscale rent lift at 5–10 km is among the largest in the sample (point estimates above +10% in the NoVA cluster), shows the negative formation effect of −3.6 to −3.8%. Iowa, a slack rural rental market with low baseline price-of-entry inferred from state-level supply-elasticity data, shows the positive formation effect of +11.8%. Texas, where the hyperscale footprint is concentrated in the mixed Austin and Dallas exurbs, falls in between and is noisy (+5.2% at 1–2 km, −9.2% at 2–5 km, neither significant).

The sign pattern would be consistent with the land-cost crowd-out mechanism if cost-of-entry differences drive the cross-state ordering. We do not read the three-cell pattern as confirming the mechanism. A more definitive cross-jurisdiction test would require state-of-incorporation panels from additional DC-hosting states (Arizona, Washington, North Carolina) and ideally an active-plus-dissolved registry like Virginia's. We have filed the relevant requests, but the data are not yet in hand.

An entity-type decomposition (`cf_231`) localizes the effect to non-passive-holding entities, consistent with genuine local activity rather than registered-agent shells, though it does not prove it. We classified each filing into one of six buckets ordered by priority: passive-holding (name matches *Holdings* / *Properties* / *Investments* / *Trust* / *Capital* / *Ventures*), foreign (out-of-state registration), nonprofit, professional (PLLC / Professional Corporation), self-managed (registered-agent address equals principal address, a sole-proprietor LLC indicator), and standard-operating (everything else). The contextual divergence is concentrated in real-economy buckets in both states.

Table 11: New business formation by entity type, hyperscale rings

State	Entity-type bucket	n records	1–2 km		2–5 km	
			detrend	p	detrend	p
VA	*Self-managed (sole-prop LLC)*	272k	−3.0%	0.028	−3.2%	0.028
VA	Standard operating	135k	−2.0%	0.13	−1.9%	0.20
VA	Passive holding	29k	−1.0%	0.25	−0.3%	0.70
VA	Foreign	5k	−0.1%	0.76	−1.0%	0.029
IA	*Standard operating*	78k	+8.4%	0.044	+5.4%	0.36
IA	Self-managed	72k	+4.9%	0.19	+1.2%	0.76
IA	Passive holding	20k	+1.4%	0.63	+6.3%	0.28
IA	Foreign	28k	−0.1%	0.73	+2.1%	0.26
IA	Nonprofit	10k	−0.6%	0.82	+0.5%	0.85
IA	Professional	3k	−0.9%	0.46	−0.1%	0.97

Notes: Same detrended ring-DiD specification as Table 10, restricted to the indicated entity-type bucket. "Self-managed" = principal address equals registered-agent address (a proxy for sole-prop LLC structure); "Passive holding" = name matches a holding/investment regex (`cf_231`); "Foreign" = registered outside the host state. VA: file is all LLCs, so no nonprofit / professional row; the self-managed indicator is built from the (principal, RA) address comparison. IA: explicit `Corporation Type` field carries the nonprofit / professional / foreign tags; self-managed is again the (principal, RA) match. Buckets are mutually exclusive and applied in the priority order listed in the text. $^{*},^{**},^{***}$: $p < 0.10, 0.05, 0.01$.

In Virginia, the aggregate −3.8% at 2–5 km is concentrated in *self-managed LLCs* (−3.2%, $p = 0.028$), sole-proprietor or single-member structures with no separate registered agent. These are the smallest, most personally exposed operators (the local handyman, the single-family-rental LLC, the side-business consulting LLC). They are also the most cost-sensitive to commercial-rent pressure and the most likely to substitute to a cheaper address farther out when rents rise in the immediate ring. The standard-operating bucket is also directionally negative (−1.9%, $p = 0.20$) but the effect is smaller. Passive holdings, foreign entities, nonprofits, and professional-service LLCs are essentially null. The negative entry effect is not a shell-LLC phenomenon.

In Iowa, the aggregate +11.8% at 1–2 km is concentrated in *standard-operating entities* (+8.4%, $p = 0.044$), real businesses with professional structure (separate registered agent, non-shell name). Self-managed and passive-holding categories are directionally positive but not significant. Foreign, nonprofit, and professional entities are null. The positive entry effect is also *not* a shell or passive-investment artifact.

The reduced entry is not just redirected to nearby rings. A natural alternative reading of the VA negative is that would-be operators simply file at a cheaper address slightly farther out, in which case the 5–10 km and 10–25 km rings should show a positive offset. They do not: VA 5–10 km detrended is −1.1% ($p = 0.20$) and 10–25 km is −1.4% ($p = 0.11$), both directionally negative. The reduced entry in the immediate ring appears to be either a net loss of would-be formations from the broader catchment, realized at locations outside the 50 km comparison radius, or not realized at all.

The national county-level BFS row is null (+1.4% detrended, $p = 0.49$), consistent with counties being too coarse to isolate the ring-level signal from surrounding null space. This pattern underscores why fine-grained ring-level data are essential to detect the sign-reversal between rural and exurban contexts.

As a robustness slice, we link each Secretary-of-State filing to the Veridion firm master by normalized name, ZIP, and state, which recovers a NAICS code for about 2 percent of records. This subset is the filings with a detectable digital footprint, web presence, and operational track record by Veridion's coverage standards. On this much-smaller subset, which is selection-biased toward established operating businesses, the contextual effects attenuate substantially. The Virginia −3.8% at 2–5 km becomes −0.6% (NS), and the Iowa +11.8% at 1–2 km becomes +0.5% (NS).

Two interpretations fit. The pessimistic read is that shell-LLC behavior drives the aggregate results without reflecting real economic activity. Our preferred read is that Veridion's universe is selection-biased toward established firms with digital footprints. By construction it excludes the population we want to measure, which is newly-forming entities responding to a local treatment. New small operators have not yet had time to develop a Veridion-visible web presence. And the entity-type decomposition above shows the response is concentrated in self-managed and standard-operating buckets, the categories most under-represented in Veridion relative to their share of SoS filings.

We therefore report the SoS aggregate as the primary measurement and the Veridion-matched subset as a complementary slice. The slice tells us the response runs through smaller, newer, not-yet-digital entities rather than through already-established operating firms.

Two interpretive caveats remain. First, the Iowa data are active-only (dissolved entities are not included), which biases the IA flow count toward formations that survived. The direction of this bias is not signed without observing exits. If DC-adjacent formations survive longer, the active-only panel could amplify the positive estimate. If omitted short-lived rural formations are more common away from DC catchments, the panel could attenuate it instead. The entity-type decomposition mitigates the concern but does not remove it. The positive effect is concentrated in standard-operating entities (the modal real-business form), not in self-managed or passive-holding shells that might inflate active-only counts.

Second, the Texas hyperscale sample is small ($n = 14$) and serves more as a directional check than a power test. North Carolina ($\sim$ 15 HS DCs) and Arizona ($\sim$ 8) would be informative intermediate-density cases. Both require paid or form-mediated access to bulk SoS data and are reserved for the revision.

The firm-entry response to a hyperscale opening is not a structural feature of the data-center treatment itself. It depends on the local density and supplier base of the host area. In rural Iowa, hyperscale arrival is followed by increased *entry* of real operating businesses (standard-operating LLCs and corporations) in the immediate 1–5 km ring. These are the kinds of entities needed to support the build and the in-migrating workforce.

In dense exurban Virginia, the same treatment is followed by reduced *entry* of sole-proprietor and small operating entities in the same ring, with no offsetting increase at nearby rings. This is consistent with rent capitalization, parcel-availability constraints, and a weak local demand response from the small DC operations workforce. The two effects cancel at the national county-level aggregate.

The contextual reading aligns with the rent and capital findings already in the matrix. Rent rises and a capital footprint appears in both contexts. But the flow of new economic actors into the immediate commercial neighborhood responds to those pressures in opposite directions, depending on whether the catchment has spare commercial supply (rural) or not (exurban).

Data Availability Statement

This study combines publicly available data with several proprietary datasets obtained under license. All code that constructs the analysis datasets from the raw sources and reproduces every table and figure will be deposited in the AEA Data and Code Repository. The deposit includes the publicly available inputs and all author-constructed files, namely the data-center registry, the operating and construction-start dates, and the announced-then-cancelled cohort. The proprietary inputs cannot be redistributed. For each, we report the provider and access terms and include the exact product identifiers, date ranges, geographic filters, and extraction code, so that a licensed user can regenerate the facility-by-ring panels and reproduce all results.

Publicly available data.

- Satellite nighttime lights: VIIRS Day/Night Band Monthly Composite (NOAA, product VCMCFG), accessed through Google Earth Engine.
- Daytime imagery for construction dating: Sentinel-2 and Landsat surface reflectance (ESA and USGS), accessed through Google Earth Engine.
- Data-center registry: base layer from the IM3 Open Source Data Center Atlas (Pacific Northwest National Laboratory), augmented by the authors from operator press releases, state economic-development announcements, the Good Jobs First subsidy tracker, trade press, DatacenterMap, Cloudscene, and data-center REIT filings. The compiled registry, dates, and cancelled cohort are in the deposit.
- Construction timing: Federal Aviation Administration Obstruction Evaluation / Airport Airspace Analysis filings (Form 7460-1).
- Employment and establishments: Quarterly Census of Employment and Wages (Bureau of Labor Statistics) and County Business Patterns (Census Bureau).
- Migration and income: IRS Statistics of Income county-to-county migration data.
- Housing supply: Census Building Permits Survey. Population: 2020 Census Centers of Population (tract population-weighted centroids).

- Workplace employment: LEHD Origin-Destination Employment Statistics (LODES, Census Bureau).
- Business formation: Census Business Formation Statistics, and state business-registration bulk files for Virginia, Texas, and Iowa (Secretaries of State and data.iowa.gov).
- Local public finance: Census of Governments and the Annual Survey of School System Finances (F-33).
- Electricity: U.S. Energy Information Administration Form 861.
- Instrument-search inputs: the InterTubes long-haul fiber map [Durairajan et al., 2015]; the 1980 county urban-college share (IPUMS NHGIS); and transmission-substation locations (Homeland Infrastructure Foundation-Level Data).

Proprietary data (available under license, not redistributable).

- Job postings: LinkUp Job Records [LinkUp, 2025] and WageScape salary postings [WageScape, 2022].
- Residential rents: RentHub Rental Data [RentHub, 2022].
- Consumer spending: SafeGraph Spend Patterns [SafeGraph, 2022].
- Foot traffic: Advan Weekly Patterns [Advan Research, 2025].
- Firmographics: Veridion firm master data.

The five Dewey-distributed datasets are available to researchers through a subscription to Dewey Data (`https://www.deweydata.io`); Veridion firmographics are available under license from Veridion (`https://veridion.com`). The authors accessed these data under institutional license and are not permitted to redistribute the raw files. The deposit provides the dataset identifiers and the extraction and aggregation code required to rebuild the analysis files from a licensed copy.

References

Advan Research. Foot traffic / weekly patterns plus [dataset]. `https://doi.org/10.82551/C103-N851`, 2025.

Fernando Alvarez, David Argente, Joyce Chow, and Diana Van Patten. Data centers and local economies in the age of AI: A shift-share approach. NBER Working Paper 35194, National Bureau of Economic Research, 2026.

Dany Bahar and Greg C. Wright. Data centers and local employment. Working paper, 2026.

Ramakrishnan Durairajan, Paul Barford, Joel Sommers, and Walter Willinger. InterTubes: A study of the US long-haul fiber-optic infrastructure. In *Proceedings of the 2015 ACM Conference on Special Interest Group on Data Communication (SIGCOMM)*, pages 565–578, 2015.

Michael Greenstone, Richard Hornbeck, and Enrico Moretti. Identifying agglomeration spillovers: Evidence from winners and losers of large plant openings. *Journal of Political Economy*, 118(3): 536–598, 2010.

LinkUp. Job records [dataset]. https://doi.org/10.82551/V091-HN53, 2025.

Carlianne Patrick. Identifying the local economic development effects of million dollar facilities. *Economic Inquiry*, 54(4):1737–1762, 2016.

Carlianne Patrick and Mark Partridge. Agglomeration spillovers and persistence: New evidence from large plant openings. Working Paper CES-22-21, U.S. Census Bureau Center for Economic Studies, 2022.

RentHub. Rental data [dataset]. https://doi.org/10.82551/5MMV-7M10, 2022.

SafeGraph. Spend patterns [dataset]. https://doi.org/10.82551/NSF5-R186, 2022.

Reid Taylor, Owen Kay, and Robert Reaser. Processing power: The effect of data centers on wholesale electricity markets. Working Paper 2606, Federal Reserve Bank of Dallas, March 2026.

Veridion. Firmographics (company core profiles) [dataset]. https://doi.org/10.82551/RTSB-1K62, 2026.

WageScape. Job postings with salary [dataset]. https://doi.org/10.82551/TSAG-NZ08, 2022.