

WORKING PAPER · DRAFT FOR REVIEW

# Welcome the Plants, Question the Abatements: The Local Economic Effects of Data Centers and the Incentives Used to Attract Them

A Two-Ledger Evaluation of Data-Center Investment and Subsidy in U.S. Counties

**Luke Hill**

American Institute for Economic Research (Summer Intern) · Hillsdale College

May 2026

## ABSTRACT

When a data center opens in a U.S. county it brings an unusual bundle: hundreds of millions of dollars of long-lived capital, large electricity and water draws, and a thin permanent payroll. This paper asks what such a facility does to local investment, the tax base, and employment, and whether the public incentives used to attract it justify their cost. The organizing move is to separate two objects that public debate routinely conflates. Ledger A is the investment itself: its capital, output, low-service-burden tax base, modest workforce, and any agglomeration spillovers. Ledger B is the firm-specific subsidy used to attract it. We bring the modern place-based-policy and heterogeneity-robust difference-in-differences toolkit to bear on a county panel of staggered first openings, anchor the employment evidence to Bahar and Wright (2026), and evaluate subsidies through a cost-per-induced-job and marginal-value-of-public-funds accounting that applies the but-for rates of Bartik (2018a). The original county-panel estimates are forthcoming and are flagged as such throughout; the conclusions drawn here rest on the external causal literature. The evidence points in opposite directions on the two ledgers. The investment is plausibly net-positive locally, concentrated in hyperscale clusters; the firm-specific subsidy is

usually value-destroying because most of its dollars are inframarginal. The policy that follows is to welcome the plants, tax their capital neutrally, and retire the discretionary abatements.

---

*JEL codes:* H25, H71, R11, R58, Q41

*Keywords:* data centers, place-based policy, business tax incentives, but-for rate, marginal value of public funds, staggered difference-in-differences

*Acknowledgment:* I thank my mentors at the American Institute for Economic Research for guidance and critical feedback; the companion R/Stata reproducibility pipeline and this paper build on my earlier work on federal data-center legislation, and the views and any errors are my own.

---

## 1. Introduction

The artificial-intelligence build-out has made the data center the most conspicuous piece of private capital formation in the contemporary American landscape. The hardware required to train and serve large models has turned an obscure category of industrial real estate into a national investment story. McKinsey (2025) projects that scaling global data-center capacity will require on the order of \$5.2 trillion for AI alone, rising to roughly \$6.7 trillion in total data-center investment, by 2030, with more than forty percent of it expected to land in the United States. That this is a forecast from a consultancy, not a measured fact, should be kept in view; but the contemporaneous data tell a consistent story. Even as total U.S. construction spending contracted modestly in 2025, outlays on data-center construction ran at an annual level near \$41 billion, up roughly a third year over year, and reached an annualized rate near \$45 billion by December.<sup>1</sup> The facilities consuming this capital are unusually heavy on it: a single hyperscale campus represents hundreds of millions of dollars of long-lived plant operated by, on most accounts, only a hundred to two hundred permanent staff (Bahar and Wright 2026).

This combination—enormous capital, large electricity and water draws, and a thin permanent payroll—has made data centers a flashpoint in state and local fiscal politics. States have competed aggressively to attract them, chiefly through sales- and use-tax exemptions on servers and equipment, property-tax abatements, and discretionary grants. The bills are now large enough to register in state budgets: Virginia is estimated to forgo on the order of \$1.6 billion in state revenue annually through its data-center exemption, and ten or more states each forgo more than \$100 million per year (Tarczynska and LeRoy 2025). Whether such concessions purchase anything for the public that would not have arrived anyway is precisely the question this fiscal politics turns on, and it is the question this paper takes up.

We ask a single, tractable question: when a data center opens in a U.S. county, what happens to local investment, the tax base, and employment—and do the public incentives used to attract these facilities justify their cost? Our answer rests on a distinction that the public debate routinely collapses but that public economics insists upon: the investment and the subsidy are separate objects of evaluation. *Ledger A* is the facility itself—its capital, output, and property-tax base, the modest permanent workforce and the local multiplier it supports (Moretti 2010), and any agglomeration spillovers to incumbent firms (Greenstone, Hornbeck, and Moretti 2010). *Ledger B* is the firm-specific subsidy used to attract it. The thesis we defend, and which the external evidence supports, is compact: welcome the plants, question the abatements. Data centers bring genuine capital deepening and a tax base that demands almost no public services—Loudoun County, Virginia, derives roughly 38 percent of its general-fund revenue from facilities occupying about 4 percent of its commercial parcels,<sup>2</sup> a near-textbook public-finance result. The correctable failure is not the investment but the subsidy, most of whose dollars are *inframarginal*—paid for activity that would have occurred regardless. The right line, in the framing of Walczak (2025), is to defend neutral non-taxation of business inputs while opposing discriminatory, firm-specific carve-outs.

Two features distinguish our approach. First, the empirical strategy. We exploit the staggered timing of first data-center openings across a U.S. county panel and estimate dynamic treatment effects with the heterogeneity-robust difference-in-differences toolkit—Callaway and Sant’Anna (2021) as the primary estimator, with the negative-weighting pathologies of naive two-way fixed effects (Goodman-Bacon 2021; de Chaisemartin and D’Haultfoeulle 2020) diagnosed and a menu of robustness estimators overlaid—paired with honest pre-trends sensitivity (Roth 2022; Rambachan and Roth 2023). Second, the cost-benefit analysis is built on two ledgers and two cost-per-job standards: the cost per *promised* job that boosters cite, and the cost per *induced* job that follows once a but-for rate near 12 percent is applied (Bartik 2018a, 2018b), expressed through the marginal-value-of-public-funds framework of Hendren and Sprung-Keyser (2020). This matters because there is, remarkably, almost no peer-reviewed economics specific to data centers; the only rigorous causal employment evidence is Bahar and Wright (2026), and the remainder of the discussion lives in gray and advocacy literature. This paper brings the modern place-based-policy and staggered-DiD apparatus to the question for the first time. The original county-panel estimates that would populate Ledger A are forthcoming; throughout, we mark clearly where an author’s own estimate will be inserted and rest the substantive conclusions on the external literature in the interim.

The remainder proceeds as follows. Section 2 supplies institutional and industry background. Section 3 develops the two-ledger conceptual framework. Section 4 reviews the literature on incentive evaluation, agglomeration, and tax competition. Section 5 describes the county-panel data and treatment definition, and Section 6 the empirical strategy. Section 7 presents the design and predicted results for investment, the tax base, and employment; Section 8 examines Loudoun County via synthetic control. Section 9 develops the incentive cost-benefit analysis, Section 10 engages the energy, water, and few-jobs counterarguments, and Section 11 draws policy implications. Section 12 concludes.

- 
1. U.S. Census Bureau, *Value of Construction Put in Place* (C30 series). The \$41 billion figure is the 2025 level; the ~\$45 billion figure is the December 2025 seasonally adjusted annual rate. ↩
  2. Loudoun County government, Data Centers public reporting; see Section 8 and the data appendix for provenance. ↩

## 2. Institutional and Industry Background

A data center is a purpose-built facility that houses the servers, storage arrays, and networking equipment on which modern computing depends, together with the power-delivery and cooling systems required to keep that hardware running continuously. Economically, the defining feature is the ratio in which it combines inputs: a data center is an unusually pure bundle of physical capital and electricity, with labor entering almost entirely during construction rather than operation. This input mix—not the novelty of the technology—is what governs the facility’s footprint on a local economy and structures the empirical questions that follow. It is also what makes the industry a clarifying test case for place-based-policy analysis, because the channels through which the investment plausibly benefits a county (capital deepening, a property and sales tax base, construction activity, possible agglomeration) are largely distinct from the channel boosters most often invoke, namely permanent jobs.

The industry is conventionally divided into two business models. *Colocation* providers lease rack space, power, and connectivity to many tenants inside a shared building; *hyperscale* operators—the large cloud and platform firms—build and run very large campuses for their own workloads. The distinction matters for local economics because scale and ownership drive the magnitude of investment and, as Section 7 documents, the presence of measurable spillovers. Bahar and Wright (2026) find that the favorable employment and wage effects of data-center entry are concentrated in hyperscale facilities and

clusters: counties that accumulate four or more facilities show roughly a 23 percent rise in information-sector employment, while isolated single colocation sites show no significant information-sector growth. Any verdict on data centers must therefore be conditioned on facility type and clustering rather than asserted for the category as a whole.

Data centers are capital-intensive because the bulk of their cost is durable equipment and the electrical and mechanical plant that supports it, and they are labor-light because, once commissioned, a fully automated facility requires only a small crew of technicians and security and facilities staff to operate. Industry staffing ranges place a typical 100-megawatt hyperscale campus at roughly 100 to 200 permanent employees, with smaller 50-megawatt sites nearer 50 to 80. The gap between construction-phase and operational employment is stark and recurs across announced projects: the Stargate campus in Abilene, Texas was paired with roughly 1,500 construction workers against 357 permanent positions, and Amazon’s New Albany, Ohio facility reported about 105 permanent jobs.<sup>1</sup> This is the empirical basis for treating data centers as a weak jobs program but a potentially strong capital and tax-base investment—a distinction that the conceptual framework in Section 3 formalizes and that the cost-benefit accounting in Section 9 makes quantitative.

Site selection is driven by the inputs that dominate the cost structure: abundant, cheap, and reliable electricity above all, then suitable land, dense fiber connectivity, water for cooling, and tax treatment. The energy demand is large in aggregate and growing quickly. The Lawrence Berkeley National Laboratory (2024) estimates that U.S. data centers consumed 176 terawatt-hours in 2023, about 4.4 percent of national electricity use and up from 58 terawatt-hours in 2014, and projects 325 to 580 terawatt-hours—6.7 to 12 percent of U.S. electricity—by 2028. The same report attributes roughly 17 billion gallons of direct cooling-water consumption in 2023, plus about 211 billion gallons consumed indirectly through electricity generation. These figures motivate the energy-, water-, and ratepayer-incidence counterarguments engaged directly in Section 10; reported in full ranges rather

than at the high end, they establish that the resource footprint is real without presuming it is decisive.

The capital outlay is correspondingly large. U.S. data-center construction spending ran at an annual level near \$41 billion in 2025, up roughly 32 percent year over year even as total U.S. construction spending fell about 1.4 percent, and reached an annualized rate near \$45 billion by December.<sup>2</sup> Looking forward, McKinsey and Company (2025) projects \$5.2 trillion for AI alone, rising to roughly \$6.7 trillion in total global data-center investment, by 2030, with more than 40 percent expected in the United States; this is a consultancy forecast with a wide uncertainty band rather than a measured outcome, and is treated as such throughout. At the state level, the Joint Legislative Audit and Review Commission of Virginia (JLARC 2024) attributes about 74,000 jobs, \$5.5 billion in labor income, and \$9.1 billion in GDP to the industry, with Virginia holding roughly 35 percent of the world's hyperscale market—a concentration the Loudoun County case study in Section 8 examines in detail. Together these magnitudes frame the central tension this paper addresses: a small permanent workforce sitting atop a very large, immobile, low-service-burden capital base.

- 
1. Project-level staffing counts are drawn from company and government project announcements rather than a single statistical source; the campus-level staffing ranges follow published industry profiles. ↩
  2. U.S. Census Bureau, *Value of Construction Put in Place* (C30 series), as in Section 1. ↩

### 3. Conceptual Framework

The analytical core of this paper is a separation that public debate over data centers routinely collapses: the distinction between the **investment** a facility

represents and the **subsidy** used to attract it. These are different objects, with different welfare properties, and conflating them is the source of most of the confusion in the policy conversation. We therefore organize the framework around two ledgers. Ledger A evaluates the facility itself—its capital stock, the output and county product it generates, the property and sales-tax base it creates, the construction and operational employment it supports, and any agglomeration spillovers to incumbent firms. Ledger B evaluates the public incentive—the sales- and use-tax exemptions, property-tax abatements, and discretionary grants offered to induce a location decision. A facility can be locally net-positive (Ledger A) while the subsidy that secured it is value-destroying (Ledger B), because the two ledgers ask different counterfactual questions. Ledger A asks what changes in a county when a data center arrives; Ledger B asks what the public outlay buys *at the margin*—how much of the investment the subsidy actually caused rather than merely paid for.

### **3.1 Spatial equilibrium and incidence: who captures the gains**

The natural setting for Ledger A is the spatial-equilibrium model of local labor markets developed by Moretti (2011) and surveyed by Glaeser and Gottlieb (2009). A new facility is a positive local labor-demand shock. In a frictionless long run with mobile workers, the incidence of that shock is shared: nominal wages rise, but in-migration and capitalization into housing costs erode part of the real-wage gain, so a portion of the benefit accrues to landowners rather than to incumbent workers. This is why the relevant question for any local-development policy is not whether activity rises but *who captures the surplus*. The same logic governs the subsidy in Ledger B. Drawing on Suarez Serrato and Zidar (2016), the incidence of a business tax break falls roughly 40 percent on firm owners, 30 to 35 percent on workers, and 25 to 30 percent on landowners. For a data center owned by a national hyperscaler, the owner share is largely captured by national shareholders, a substantial fraction of them out of region—so a county financing an abatement is, to a first approximation, transferring a meaningful part of the foregone revenue beyond its own borders. Incidence is

thus not a footnote; it is decisive for whether a locally financed incentive can plausibly benefit local residents.

### **3.2 Local multipliers applied to a labor-light technology**

Ledger A's employment channel must be handled with unusual honesty for this industry. Moretti (2010) estimates a local multiplier of roughly 1.6 non-tradable jobs per additional tradable-sector job, larger for skilled and high-technology work. The temptation is to apply this multiplier to a data center's headline investment and report a large induced-jobs figure. The discipline the framework imposes is that the multiplier operates on the *permanent operational* base, which for these facilities is very small—on the order of a hundred or two hundred staff for a large campus—not on the transient construction workforce. A modest permanent headcount, even multiplied, yields a modest permanent jobs footprint. The credible employment case for data centers is therefore not a direct-jobs case but an agglomeration case, in the tradition of Greenstone, Hornbeck, and Moretti (2010), who find incumbent-plant productivity roughly 12 percent higher in counties that won a large plant than in narrowly defeated runner-up counties. Such spillovers are real but heterogeneous, concentrated where facilities share labor and technology pools—consistent with the data-center-specific finding that employment gains appear in hyperscale clusters rather than at isolated colocation sites (Bahar and Wright 2026). Critically, Kline and Moretti (2014b) caution that local gains may be partly offset by losses elsewhere, so a measured county-level effect is not automatically a national gain.

### **3.3 The but-for problem and the inframarginal subsidy**

Ledger B turns on a single parameter: the *but-for rate*, the share of subsidized investment that would not have occurred absent the incentive. The literature is unambiguous that this share is low. Bartik (2018a) reviews 34 estimates and finds a central but-for rate near 12 percent, with studies free of obvious bias clustering lower still. The implication is that most subsidy dollars are *inframarginal*—they pay for investment that would have located in the

jurisdiction anyway, functioning as a pure transfer rather than an inducement. This is why the appropriate cost metric is the cost per *induced* job, not the cost per *promised* job: Bartik (2018b) estimates that typical incentives cost on the order of \$200,000 per job actually induced. Slattery and Zidar (2020) document the scale of the practice—discretionary deals averaging roughly \$178 million for about 1,500 promised jobs—and find little evidence that such spending raises broad growth. The inframarginal problem is sharpened by tax competition: Wilson (1999) formalizes the race-to-the-bottom, Chirinko and Wilson (2008) show state investment incentives largely relocate activity across state lines rather than create it, and Slattery (2025) estimates that subsidy competition transferred roughly \$40 billion to firms for about \$13 billion in gains, with banning subsidies lowering welfare by under 5 percent. These pecuniary and fiscal externalities mean that what is individually rational for a competing county is collectively close to zero-sum.

### **3.4 The MVPF as the evaluation lens**

To adjudicate Ledger B we adopt the Marginal Value of Public Funds of Hendren and Sprung-Keyser (2020): the ratio of beneficiaries' willingness-to-pay to the net cost to government. For a firm-specific data-center subsidy, the numerator is dominated by the transfer captured by the firm (with incidence as above), plus any capitalized wage and land gains from *induced* activity; the denominator is the outlay plus added public-service and infrastructure costs and any ratepayer cost-shift, net of new tax revenue from induced activity only. Because most dollars are inframarginal, the policy-relevant MVPF is low unless the but-for rate is high and spillovers large—a stark contrast to the MVPFs above five that Hendren and Sprung-Keyser report for children's-investment policies. The operational test we carry into the empirical sections is to solve for the *break-even but-for rate* at which the MVPF equals one; if that threshold exceeds the literature's central estimate of roughly 12 percent, the subsidy fails on its own terms.

*[Original break-even but-for calibration forthcoming from the county panel and Subsidy Tracker ledger; the framework's expected result, given Bartik's (2018a) ~12 percent central rate and the inframarginal-transfer logic of Slattery (2025), is a break-even rate well above 12 percent—i.e., a representative blanket data-center abatement with an MVPF below one.]*

## 4. Literature Review

This paper sits at the intersection of two literatures that the public debate over data centers tends to conflate: the evaluation of business incentives and the economics of large plant openings. Keeping them analytically distinct is the organizing premise of the paper, and the scholarship reviewed here supplies both the methods and the priors for doing so.

**Incentive evaluation and the but-for problem.** The framing reference is Slattery and Zidar (2020), whose survey establishes the scale and the central puzzle of state and local business incentives: the average discretionary deal runs roughly \$178 million for about 1,500 promised jobs, states spend between \$5 and \$216 per capita (on the order of 40 percent of corporate tax revenue for the typical state), and yet the broad-growth evidence is thin. The reason is the but-for problem. A subsidy creates value only for the marginal firm that would not otherwise have located in the jurisdiction; dollars paid to inframarginal firms are pure transfers. Bartik's body of work operationalizes this distinction. He documents on the order of \$50 billion per year in U.S. incentives and argues that most are too large and poorly targeted, while conceding that well-designed, distressed-area incentives can pass a cost-benefit test (Bartik 2019). His review of the empirical literature places the central but-for rate near 12 percent, with a plausible range of roughly 2 to 25 percent (Bartik 2018a), and his benefit-cost model implies that a typical incentive costs about \$200,000 per job actually induced—rather than the far smaller figure boosters obtain by dividing outlays by promised jobs (Bartik 2018b). The companion data resource, the Panel Database on Incentives and Taxes (Bartik 2017), allows incentive generosity to

be benchmarked across industries, which is useful for situating data-center carve-outs relative to manufacturing and other recipients. This strand motivates the paper's cost-per-induced-job ledger and its insistence on attaching a but-for rate to every per-job claim.

**Agglomeration spillovers from large plant openings.** The methodological lodestar is Greenstone, Hornbeck, and Moretti (2010), who use counties that narrowly lost a large plant siting as the counterfactual for counties that won it. They find that the total factor productivity of incumbent plants is about 12 percent higher five years after a winning opening, with effects concentrated where the entrant shares labor and technology pools with existing establishments. This winner-versus-loser design is the cleanest analogue to a data-center event study, and it supplies an indispensable caveat: agglomeration spillovers are heterogeneous and conditional, not automatic. The same lesson appears in the broader agglomeration literature reviewed by Glaeser and Gottlieb (2009), who emphasize that agglomeration economies are real but difficult to measure and that idea flows and density matter more than mere co-location—relevant to whether labor-light data centers, which employ few on-site workers, can plausibly generate the knowledge spillovers that justify a productivity externality at all.

**Local multipliers and local labor markets.** Moretti (2010) provides the multiplier that translates a small permanent workforce into a defensible induced-jobs range: each tradable-sector job supports roughly 1.6 additional local non-tradable jobs, with a larger multiplier for skilled and high-tech employment. His Handbook chapter (Moretti 2011) supplies the spatial-equilibrium machinery in which nominal local wage gains are partly offset by cost-of-living adjustments, so that local effects must be read in general equilibrium rather than taken at face value. Kline and Moretti (2014b), studying a century of the Tennessee Valley Authority, deliver the crucial national-accounting caveat: the program produced lasting local agglomeration gains, but those gains were largely offset by losses elsewhere, leaving modest net national benefit. For this paper the multiplier disciplines the jobs claim in

both directions—it permits crediting a hyperscale campus’s hundred-odd operating staff with some induced employment while making transparent how small the base is relative to the construction phase.

**Place-based-policy surveys and the welfare framework.** The authoritative survey is Neumark and Simpson (2015), whose reading of the enterprise-zone evidence is mixed at best, with stronger support for public-goods-type investment than for firm-specific cash. Glaeser and Gottlieb (2008) reach a compatible and pointedly skeptical conclusion: because agglomeration benefits are nonlinear, spatial subsidies are “as likely to reduce as to increase welfare,” and most large programs show little impact. Kline and Moretti (2014a) provide the welfare scaffolding, showing that whether a local subsidy raises or merely redistributes welfare hinges on the strength of agglomeration externalities and the elasticity of labor mobility. To keep the paper from reading as pre-committed, the review includes the strongest contrary evidence: Busso, Gregory, and Kline (2013) find that federal Round-I Empowerment Zones substantially raised local employment and wages without large population inflows or cost-of-living increases, at modest efficiency cost—a demonstration that a well-targeted place-based policy can work, and a standard against which open-ended data-center exemptions can be judged. The origin point of this entire “who benefits” tradition, Bartik (1991), is worth stating plainly: local job growth can produce lasting benefits, but whether it yields a net fiscal gain depends on design and financing—exactly the conditional posture this paper adopts.

**Tax competition and incidence.** Wilson (1999) is the canonical survey of the race-to-the-bottom logic by which jurisdictions wastefully compete for mobile capital. Slattery (2025) quantifies it within an auction model: banning subsidy competition would lower welfare by less than 5 percent because states compete away the surplus, transferring roughly \$40 billion to firms for about \$13 billion in gains. Chirinko and Wilson (2008) show that state investment incentives largely shift activity across state lines rather than create it nationally, so that from a national-welfare standpoint subsidy wars are close to zero-sum.

Suarez Serrato and Zidar (2016) supply the incidence that sharpens the critique: firm owners bear roughly 40 percent of a state corporate tax change, workers 30 to 35 percent, and landowners 25 to 30 percent—implying that a sizable share of any data-center tax break accrues to (often out-of-region) shareholders rather than to local residents. The welfare metric that ties the cost side together is the Marginal Value of Public Funds of Hendren and Sprung-Keyser (2020): beneficiaries' willingness-to-pay over net government cost. For a transfer-to-firm subsidy the numerator is dominated by the firm's captured rent, so the policy-relevant MVPF is low unless the but-for rate is high and spillovers large—a sharp contrast with the MVPFs above five they document for children's-investment policies.

**The data-center evidence gap.** There is, as yet, essentially no peer-reviewed economics literature on data centers specifically. The only rigorous causal evidence is Bahar and Wright (2026), whose synthetic-control analysis finds that a county's first large data center raises total private employment by 4 to 5 percent and wages by 3 to 4 percent over five to six years, with no significant home-price effect—but only for hyperscale facilities and clusters (a 23 percent rise in information-sector employment in counties with four or more facilities, versus no significant effect for single colocation sites), and with naive estimates overstating the effect by roughly threefold. Everything else specific to the industry is gray or advocacy literature—most prominently the Good Jobs First reports (Tarczynska and LeRoy 2016, 2025)—which is the best available source on subsidy magnitudes and the cost-per-job framing but must be labeled as such and cross-checked against state audits such as the Georgia evaluation (Carl Vinson Institute of Government 2025) and JLARC's Virginia report (JLARC 2024). This paper's contribution is to bring the modern place-based-policy and heterogeneity-robust difference-in-differences toolkits to bear on data centers directly, and to replace the cost-per-promised-job number with an explicit cost-per-induced-job and MVPF ledger.

## 5. Data

The unit of analysis is the county-year. We assemble a panel of the roughly 3,100 U.S. counties and county-equivalents over 2012–2023. The window is chosen so that the treatment and the principal BEA, BLS, and Census outcome series jointly support a five-year pre-period and a multi-year post-period for cohorts treated through the late 2010s; the county-level electricity covariate, whose source begins in 2016, enters only over its available years and does not bind the outcome window. The merge key throughout is `fips5`, the five-digit state-plus-county FIPS code stored as a *string with leading zeros*; we standardize it once and join every file 1:1 on county and year. This discipline is not cosmetic. Numeric storage silently drops the leading zero (01001 becomes 1001), and unharmonized boundary changes corrupt merges; we therefore reconcile codes to a single vintage using the Census Bureau’s log of substantial county changes (e.g., Shannon County, SD, 46113 to Oglala Lakota, 46102; the 2022 Connecticut transition to planning regions) and document every recode in a data appendix.

**Treatment.** A county is treated in the year its first large data center begins operating; we focus on the first opening to define an absorbing event and, separately, exploit the cumulative count and capacity of facilities for the dose-response analysis. The facility universe is the DOE/PNNL Integrated Multisector Multiscale Modeling (IM<sup>3</sup>) Open-Source Data Center Atlas, a national-laboratory product that geolocates U.S. data centers and maps each to a county FIPS. Its provenance and open license make it the reproducible backbone of the treatment list, and the synthetic-control design for the county case study follows the approach that yields the only rigorous data-center-specific causal estimates to date (Bahar and Wright 2026). The Atlas is, however, a cross-section: it records where facilities are, not *when* they opened, and it omits megawatt capacity and operational status. We therefore supply the event date,  $t = 0$ , by triangulating three independent proxies, following the methodology in Section 6. The first is plant-entry timing from the EIA Form-860 generator and plant files, geocoded to the county. The second is the first

award year in the Good Jobs First Subsidy Tracker for a facility geocoded to the county. The third is a structural break in the establishment count for NAICS 518210 in the QCEW/CBP series. As a supplementary check we use the timestamped building-edit history of OpenStreetMap footprints, where the first appearance of a tagged data-center building approximates a construction or first-mapped date. Because each proxy carries timing error, we report all results across the dating rules and treat their disagreement as a bounded source of measurement error rather than concealing it. Capacity and operational status, absent from IM3, are cross-checked against the Baxtel facility database to construct the megawatt intensity used in the dose-response specification.

**Outcomes.** The investment and output channel is captured by BEA county GDP (the CAGDP tables), the cleanest test of whether a capital-intensive, low-headcount facility raises local value added, and by the valuation of authorized building permits from the Census Building Permits Survey as an investment and tax-base proxy. The employment and wage channel relies on the BLS Quarterly Census of Employment and Wages (QCEW), an administrative near-census of UI-covered jobs reporting county-by-NAICS employment, establishment counts, and average wages; we use total private employment and wages as the primary outcomes and the data-center industry, NAICS 518210, as a caveated secondary measure. We cross-check establishment counts against the Census County Business Patterns (CBP) and, for small and rural counties where suppression bites hardest, against the Eckert–Fort–Schott–Yang imputed CBP panel. The fiscal channel draws on county personal income (BEA CAINC) and IRS county adjusted gross income. We note that BEA can no longer supply county employment—the relevant CAINC4 lines were removed and table CAEMP25 was discontinued in November 2024—so BEA serves the income and GDP channels only, with QCEW and CBP carrying employment.

**Cost side and externalities.** Subsidy dollars and promised jobs come from the Good Jobs First Subsidy Tracker, filtered to data-center deals; because the Tracker lacks a county field, we geocode each City-plus-State record to a county FIPS before aggregating to the county-year, and we cross-check awarded

amounts against realized forgone revenue in state tax-expenditure reports and GASB 77 disclosures, treating Tracker figures as source-cited lower bounds. The county-level electricity series is the NREL county hourly load file (OEDI submission 8562), an hourly megawatt record for 2016–2023 that we preprocess to annual county megawatt-hours; the EIA state-by-sector retail price and sales series, with utility service-territory matching via EIA Form-861, serves as a state-level cross-check, and the Lawrence Berkeley National Laboratory (2024) national report is the magnitude benchmark against which we sanity-check load attribution. County water withdrawals, including the thermoelectric category used to attribute the water cost of data-center electricity, come from the USGS Water Use compilation. Controls and matching covariates are the USDA Rural-Urban Continuum Codes (RUCC), OMB/Census core-based statistical area delineations, and the American Community Survey five-year county estimates (baseline population, density, income, education, age structure).

**NAICS 518210 limitations.** Two features of the industry code constrain its use. First, it undercounts: captive hyperscale facilities are frequently classified under a parent firm’s primary industry rather than under data processing and hosting, and the 2022 NAICS revision altered the 518/519 hosting definitions, so we verify code continuity across vintages. Second, because few counties host many 518210 establishments, the cells are routinely disclosure-suppressed (QCEW `disclosure_code == 'N'`; CBP noise flags). We treat suppressed and noise-flagged near-zero cells as missing, never as true zeros, and we report the suppression rate for 518210 explicitly. These limitations are precisely why total county employment and county GDP—not the narrow industry code—serve as the primary outcomes, with 518210 reserved as a secondary, honestly-caveated measure.

## 6. Empirical Strategy

Our object of inference is the dynamic causal effect of a county's first large data center on local investment, the tax base, and employment. The setting is a county-FIPS panel in which facilities open in different years, so treatment is *staggered*: a county is assigned a cohort  $E_i$  equal to the calendar year its first qualifying facility comes online, and never-treated counties have  $E_i$  missing. This is precisely the design in which the workhorse two-way fixed-effects (TWFE) event-study regression is no longer trustworthy, and the bulk of this section explains the estimator we use instead and the diagnostics that justify it.

### 6.1 Why not two-way fixed effects

The intuitive specification regresses a log outcome on county and year fixed effects and a treatment indicator (or a set of relative-time leads and lags). Under heterogeneous and dynamic treatment effects—exactly what we expect, since a hyperscale cluster and a single colocation site should not move a county identically—this estimator is biased. Goodman-Bacon (2021) shows the TWFE coefficient is a variance-weighted average of all possible  $2 \times 2$  difference-in-differences comparisons embedded in the panel, and that one family of those comparisons uses *already-treated* counties as the control group for *later-treated* ones. When the early adopters' effects are still evolving, their post-period trend enters the comparison with the wrong sign. De Chaisemartin and D'Haultfoeuille (2020) make the point sharply: the implied weights can be negative, so the TWFE coefficient can be negative even if every county-period treatment effect is positive. Their survey (de Chaisemartin and D'Haultfoeuille 2023) documents how pervasive the naive design remains in applied work, which is why we treat it as a benchmark to be diagnosed rather than as our estimate. We make the diagnosis explicit by running the Goodman-Bacon decomposition on the naive TWFE specification and reporting the share of identifying weight that rests on the forbidden already-treated-as-control comparisons; a large share is itself evidence that the naive number should be discounted.

## 6.2 Primary estimator: Callaway and Sant’Anna

Our headline estimates come from the Callaway and Sant’Anna (2021) group-time average treatment effect on the treated,  $ATT(g, t)$ , which is the effect of opening for cohort  $g$  observed in period  $t$ . Each  $ATT(g, t)$  is identified by comparing the change in the outcome for cohort  $g$ , from the period before treatment to  $t$ , against the contemporaneous change among counties not yet treated by  $t$ —a control group that, by construction, never includes already-treated units, eliminating the contamination above. We use the not-yet-treated comparison group so that late adopters and (where available) never-treated counties both inform the counterfactual, and we estimate each cell with the doubly-robust method, which combines an outcome-regression and an inverse-probability-weighting model so that consistency requires only one of the two to be correctly specified. County covariates that plausibly drive both siting and local growth—baseline population, log electricity price, and manufacturing and broadband shares—enter through this doubly-robust step rather than as additional fixed effects, which would reintroduce the bad comparisons. We then aggregate the matrix of  $ATT(g, t)$  two ways: into an event study indexed to time-since-opening, our main graphical object, and into a single overall ATT, our one-number summary of “what a data center does to the county.” *[Original estimates forthcoming from the county panel; expected sign and magnitude are based on Bahar and Wright (2026)—roughly a 4–5 percent rise in private employment over five to six years, concentrated in hyperscale clusters—and on Greenstone, Hornbeck, and Moretti (2010) for the spillover channel.]*

## 6.3 Specification

Outcomes enter in logs or per capita: log QCEW and CBP employment in NAICS 518210 and in total private employment, log building-permit valuation and assessed value as the investment and tax-base proxies, log establishments, and log county GDP. The merge key is `fips5`, a five-digit string with leading zeros, joined 1:1 on county and year (a numeric FIPS silently drops the leading zero and corrupts the merge). Standard errors are clustered at the county level, the unit of treatment assignment. The event window runs from  $t = -5$  to  $t = +5$

with the endpoints binned, and we omit  $t = -1$  as the reference period; the 2012–2023 panel described in Section 5 is wide enough to support this window for the cohorts that anchor identification, with later-treated cohorts contributing the shorter horizons they can support. Because counties frequently gain *multiple* facilities, we also estimate a dose-response specification using cumulative installed capacity (megawatts) or facility count as a non-binary, non-absorbing treatment.

#### **6.4 Robustness menu**

We overlay four alternative estimators on a single event-study plot, since convergence across methods with different assumptions is the real credibility argument. The Borusyak, Jaravel, and Spiess (2024) imputation estimator is the most efficient and supplies a clean pre-trend test; the Sun and Abraham (2021) interaction-weighted estimator is the closest robust analogue to the familiar event-study regression and the easiest to communicate; the de Chaisemartin and D’Haultfoeuille (2024) dynamic estimator handles the non-binary, non-absorbing capacity treatment that the others cannot; and the naive TWFE estimate appears as the “what the old method would have said” line. Bahar and Wright’s (2026) finding that naive estimates overstate the effect roughly threefold is the empirical reason this comparison matters.

#### **6.5 Honest pre-trends**

A flat, insignificant pre-period is not proof of parallel trends. Roth (2022) shows pre-tests are often underpowered against exactly the trends that would most bias the estimate, and that conditioning on having “passed” a pre-test can worsen coverage. We therefore pair the visual pre-trend plot with the Rambachan and Roth (2023) honest difference-in-differences sensitivity analysis, reporting the relative-magnitude breakdown value  $\bar{M}$ —how large a post-period violation of parallel trends, relative to the worst pre-period violation, the data tolerate before the effect loses significance. This directly confronts the natural objection that data centers simply land in counties that were already booming.

The full pipeline is implemented in Stata (`csdid/drddid`, `did_imputation`, `eventstudyinteract`, `did_multiplegt_dyn`, `reghdfe`, `bacondecomp`, `honestdid`) with free R equivalents (`did`, `didimputation`, `fixest::sunab`, `DIDmultiplegtDYN`, `bacondecomp`, `HonestDiD`); because Stata is not installed locally, the R track is the reproducible default.

## 7. Results: Investment, Tax Base, and Employment

This section reports the design and expected findings for the three outcome families in Ledger A—local investment, the tax base, and employment—estimated on the staggered county panel described in Sections 5 and 6. Because the author’s own event-study estimates are not yet finalized, the discussion below states the design, the predicted signs, and the magnitudes implied by the external causal evidence, and flags each place where an original estimate will be inserted. The interpretive spine is consistent throughout: a data center is a large, durable capital deployment with a small permanent payroll, so the strongest and most precisely estimable signal should appear in the investment and output channels, a clean and economically large signal in the tax base, and a real but quantitatively modest signal in employment that is concentrated in hyperscale clusters rather than in single colocation sites.

### 7.1 Reading the event studies

Each outcome is presented as a Callaway and Sant’Anna (2021) dynamic event study, with the overall aggregated ATT reported as the headline “what a data center does to a county” number, and with the Borusyak, Jaravel, and Spiess (2024), Sun and Abraham (2021), de Chaisemartin and D’Haultfoeuille (2024), and naive two-way fixed-effects estimates overlaid for comparison. The discipline of Section 6 carries through here: an insignificant pre-period is not read as proof of parallel trends (Roth 2022), so each headline post-treatment effect is paired with the Rambachan and Roth (2023) relative-magnitude breakdown value  $\bar{M}$ , reported as the share of pre-trend variation that the estimate can tolerate before losing significance. The single most consequential

benchmark for interpretation is Bahar and Wright’s (2026) finding that naive estimators overstate the data-center employment effect by roughly a factor of three; the gap between our overlaid naive TWFE line and the heterogeneity-robust estimators is therefore not a nuisance but a reported result, and the Goodman-Bacon (2021) decomposition documents how much of the naive coefficient rests on forbidden already-treated-as-control comparisons.

## **7.2 Investment and output: the strongest expected signal**

The investment channel is where the case for data centers is most defensible, precisely because it does not depend on a large headcount. Using log county GDP (BEA CAGDP) and the log of total building-permit valuation as the investment proxy, the prediction is a sharp, well-identified positive jump beginning at or just before  $t = 0$ —coincident with the construction phase—that partially recedes but settles at a permanently higher level once the facility is energized and capitalized into the local capital stock. This pattern reflects the labor-light, capital-heavy production technology documented in Section 2 and is the local micro-counterpart to the national capital-deepening that McKinsey (2025) projects, a forecast we present with its wide uncertainty band rather than as a measured fact.

*[Original estimates forthcoming from the county panel; expected sign/magnitude based on Bahar and Wright (2026): a positive and statistically significant overall ATT on log county GDP, with a transitory construction-phase spike in building-permit valuation followed by a persistent level shift. The investment effect is expected to be the most precisely estimated of the three outcome families because it is least affected by NAICS 518210 suppression.]*

## **7.3 Tax base: a large effect with little service burden**

The tax-base result is conceptually the cleanest public-finance prediction and is developed in full for Loudoun County in Section 8. In the panel, the prediction is a large and persistent increase in assessed value and property-tax-relevant

valuation with no offsetting increase in service-cost drivers, because data centers add essentially no residents and no schoolchildren. This is the channel that most directly substantiates the “welcome the plants” half of the thesis, and it is the one where the gap between a data center and a conventional manufacturing plant of comparable assessed value is widest. The honest qualifier, carried from Section 6, is that the favorable tax-base reading concerns the *investment*, not the *abatement*: where the assessed value is sheltered by a firm-specific exemption, the realized local revenue is the post-abatement stream, evaluated against the inframarginal-subsidy benchmark of Ledger B (Section 9).

*[Original estimates forthcoming: a positive, persistent ATT on log assessed value / building-permit valuation, expected larger in proportional terms than the GDP effect and substantially larger than the employment effect—the quantitative signature of a capital-intensive, service-light land use.]*

#### **7.4 Employment: real but modest, and concentrated**

Employment is where rigor most disciplines enthusiasm. Two facts must be held together rather than traded off (Section 10): on-site permanent headcount is small—on the order of 100–200 staff for a 100-megawatt hyperscale campus—yet county-level employment can still rise meaningfully through construction activity, local multipliers (Moretti 2010), and agglomeration spillovers of the kind Greenstone, Hornbeck, and Moretti (2010) identify for large plant openings. Our anchor is Bahar and Wright (2026): a county’s first large data center raises total private employment by roughly 4–5 percent and wages by 3–4 percent over five to six years, with no significant home-price effect. Crucially, these are not uniform: information-sector employment rises by about 23 percent in counties with four or more facilities, while single colocation sites show no significant information-sector growth. We therefore report employment heterogeneously, splitting hyperscale from colocation and estimating a dose-response in cumulative capacity (megawatts) or facility count via de Chaisemartin and D’Haultfoeuille (2024). We also separate the

transitory construction boom from the permanent operational footprint, reporting the impact and post-impact event-time coefficients distinctly so that a large  $t = 0$  employment spike is not mistaken for a durable jobs gain.

*[Original estimates forthcoming: an overall private-employment ATT expected in the neighborhood of the Bahar and Wright (2026) 4–5 percent range but plausibly smaller and noisier given NAICS 518210 suppression; significant information-sector growth expected only in the four-or-more-facility cluster stratum, with the naive TWFE overlay expected to exceed the robust estimates by roughly threefold, consistent with Bahar and Wright (2026).]*

The synthesis these results are built to support is that a data center is a strong capital and tax-base proposition and, at best, a modest and conditional jobs proposition—favorable on net for the local economy, but only once the construction boom is distinguished from the permanent footprint and only where clustering generates the spillovers the single-site data do not show.

## **8. Case Study: Loudoun County, Virginia**

The county-panel estimates in Section 7 establish an average effect. A single county can make the mechanism concrete, and Loudoun County, Virginia —“Data Center Alley,” the densest concentration of data-center capacity on earth—is the natural limiting illustration of Ledger A. The county hosts the eastern anchor of an industry that, statewide, accounts for roughly 74,000 jobs, \$5.5 billion in labor income, and \$9.1 billion in Virginia GDP, and that places the Commonwealth at roughly 35 percent of the world’s hyperscale market (JLARC 2024). What makes Loudoun valuable for our argument is not the headline aggregates but the fiscal anatomy beneath them: the clean separation of a large tax base from a negligible public-service burden.

The central fact is a striking disproportion. Data centers occupy on the order of 4 percent of the county’s commercial parcels yet generate roughly 38 percent of

its general-fund revenue, contributing on the order of \$100 million in new revenue per year.<sup>1</sup> This is close to a textbook public-finance result. The facilities house servers, not households: they enroll no schoolchildren, generate little traffic, and demand almost none of the social services that drive the marginal cost of local government. The taxed object—high-value computing and electrical equipment subject to the county’s business tangible personal property levy—is precisely the kind of mobile, high-assessed-value capital that Suarez Serrato and Zidar (2016) study, but here it is taxed locally rather than abated. The fiscal consequence has been a decade of real-property-tax-rate reductions, from \$1.145 per \$100 of assessed value in 2016 to \$0.805 in 2025—a cut in every intervening year, leaving Loudoun with the lowest rate in Northern Virginia. As an accounting matter, the data-center base has substantially offset the levy on resident homeowners: the revenue incidence falls disproportionately on the immobile commercial capital rather than on residents, which is the mirror image of the out-of-region owner share that makes the *subsidy* objectionable in Section 9.

This is the favorable side of the ledger stated honestly, and it should be read with two cautions. First, it is a description of revenue incidence within one jurisdiction, not a national-welfare claim; to the extent the capital would otherwise have located elsewhere in the United States, the local gain is partly a transfer (Kline and Moretti 2014b; Chirinko and Wilson 2008). Second, the result is conditional on the cluster: Loudoun is the polar case of the hyperscale agglomeration that Bahar and Wright (2026) find drives essentially all of the measured employment response, not the single-facility colocation site where they find none.

Identifying the causal contribution of the cluster, rather than merely describing Loudoun’s prosperity, requires a credible counterfactual for a single treated unit. We construct it with the synthetic-control method of Abadie, Diamond, and Hainmueller (2010), forming “synthetic Loudoun” as a convex combination of donor counties—drawn from the never-treated and not-yet-treated pool—matched on pre-period assessed value per capita, employment,

population, and income. Because Data Center Alley’s build-out long predates a clean single break and pre-period fit for so distinctive a county is necessarily imperfect, we prefer as the headline estimator the synthetic difference-in-differences of Arkhangelsky et al. (2021), which augments the unit weights with DiD time weights and is more robust both to imperfect pre-treatment fit than pure synthetic control and to non-parallel trends than pure DiD. Inference proceeds by placebo permutation across donor counties, and we report the full gap path rather than a single post-period number.

*[Original synthetic-control and synthetic-DiD estimates for Loudoun’s assessed tax base, GDP, and employment forthcoming; expected sign and magnitude—a large, growing post-period gap in assessed value and a more modest employment gap consistent with the cluster effect—based on Bahar and Wright (2026) and the county fiscal record described above.]* The descriptive disproportion is not itself the causal estimate, but it bounds the prize: a tax base of this scale, attached to so light a service footprint, is the strongest version of the case for welcoming the plants—and, as Section 9 argues, the strongest reason to doubt that any abatement was needed to attract them.

- 
1. Loudoun County government, Data Centers public reporting and adopted budget documents; figures are official county statements rather than an independent estimate, and the causal contribution is the object of the synthetic-control exercise described below. ↩

## **9. Incentive Cost-Benefit Analysis**

This section turns the conceptual two-ledger framework of Section 3 into an explicit accounting. Ledger A—the investment itself—has already been evaluated through the design and predicted estimates of Sections 7 and 8. The question here concerns Ledger B: the firm-specific subsidy used to attract the

facility. The central claim is that the public conversation evaluates these subsidies against the wrong benchmark, and that against the right one most data-center incentives are predicted to fail their own cost-benefit test.

### 9.1 Two cost-per-job standards

Boosters and many fiscal-impact studies report the *cost per promised job*: the disclosed subsidy value divided by the number of operational jobs the deal is contracted to deliver. By this metric data centers already look expensive. Good Jobs First's profile of eleven megadeals exceeding \$2 billion in incentives put the average at roughly \$1.95 million per permanent job, with an outlier near \$6.4 million for Apple's North Carolina facility (Tarczynska and LeRoy 2016). Its 2025 update reports Illinois spending \$5.4 billion to support 339 data-center jobs—about \$1.4 million each (Tarczynska and LeRoy 2025). For comparison, the average discretionary deal across all industries is roughly \$178 million for about 1,500 promised jobs, or near \$120,000 per job (Slattery and Zidar 2020); data-center deals are an order of magnitude more expensive per head.

But cost per promised job is the wrong denominator, because it credits the subsidy with every job—including jobs that would have appeared without it. The welfare-relevant denominator is the number of jobs *actually induced*: promised jobs multiplied by the but-for rate  $b$ , the share of location decisions genuinely tipped by the incentive. Reviewing thirty-four estimates from thirty studies, Bartik (2018a) finds a median but-for of about 12.7 percent, with the lower-bias studies clustering near 3.4 percent and a plausible range of 2 to 25 percent. At  $b \approx 0.12$ , dividing by the induced rather than promised count multiplies the apparent cost roughly eightfold. By the paper's own arithmetic, the \$1.95 million-per-promised-job figure becomes on the order of \$16 million per *induced* job, and the \$1.4 million Illinois figure approaches \$12 million. Even at a generous  $b = 0.30$  the implied cost per induced job exceeds \$6 million—still far above Bartik's (2018b) benchmark that typical business tax incentives run about \$200,000 per induced job, and that customized public

services (infrastructure, training) cost roughly a third as much. Table 9.1 reports the full sensitivity grid for  $b \in \{0.05, 0.12, 0.30\}$ . The reason data centers fare so poorly on this standard is structural rather than accidental: as Section 2 established, the facilities are capital-intensive and labor-light, so the denominator is small whatever the but-for rate.

## 9.2 An MVPF for a representative subsidy

The cost-per-induced-job ratio captures labor outcomes but ignores who actually pockets the transfer and the fiscal side of the ledger. The marginal value of public funds (MVPF) of Hendren and Sprung-Keyser (2020) supplies a unified metric: the ratio of beneficiaries' willingness to pay to the net cost to government. For a firm-specific carve-out the numerator accrues mostly to the firm rather than to local households, because the subsidy is, in the inframarginal share, a pure transfer. Incidence estimates put roughly 40 percent of a state corporate tax change on firm owners, with the remainder split between workers and landowners (Suarez Serrato and Zidar 2016)—and for hyperscale operators those owners are largely national shareholders, a substantial share of them out of region, so much of the local fisc's outlay leaves the community. The numerator is therefore the firm's captured transfer plus the induced share of wage and land capitalization.

The denominator is the net government cost: the subsidy outlay, plus the added public-service and infrastructure cost, plus any ratepayer cost-shift documented by Martin and Peskoe (2025), minus the new tax revenue generated by *induced* activity only. Inframarginal revenue does not net against the cost, because that activity—and its taxes—would have arrived regardless. With most of the willingness-to-pay flowing to non-local owners and only the induced fraction of revenue offsetting the outlay, the framework predicts an MVPF for a representative blanket data-center subsidy that is well below one. *[Original MVPF estimate forthcoming from the county panel and Subsidy Tracker ledger; based on Suarez Serrato and Zidar (2016) incidence, Bartik (2018a) but-for, and the induced effects of Bahar and Wright (2026), the*

*expected value is materially below 1.0.]* The contrast with Hendren and Sprung-Keyser's (2020) catalog is stark: investments in children's health and education routinely return MVPFs above five, and several are effectively infinite, because they pay for themselves through later tax receipts. A subsidy that transfers public dollars to distant shareholders would sit at the opposite pole of that distribution.

### **9.3 The break-even but-for rate**

The most compact way to state the verdict is to solve for the break-even but-for rate  $b^*$ —the value at which the subsidy's induced fiscal and welfare benefits just equal its net cost (MVPF = 1, or equivalently NPV = 0). The logic is transparent: because the cost is fixed and only the induced fraction of activity counts as benefit,  $b^*$  is the share of facilities that must be genuinely marginal for the program to break even. *[The break-even rate will be computed from the estimated induced tax base and the realized subsidy schedule; the framework predicts  $b^*$  well in excess of 0.12.]* If  $b^*$  exceeds the central but-for of roughly 12 percent (Bartik 2018a), the subsidy fails on its own terms: it would have to tip a larger share of decisions than the literature finds plausible. Given the size of data-center capital commitments relative to the marginal value of a state tax exemption, the framework's prediction is that most of these facilities are inframarginal—they locate where power, fiber, and land are available, and the abatement is a windfall on top.

### **9.4 The abatement-NPV channel**

One genuine qualification cuts in the subsidy's favor and deserves separate treatment. Property-tax abatements are typically time-limited; after they expire the facility remains on the rolls and pays at the full rate for decades. The fair object of evaluation is therefore the net present value of the post-abatement payment stream against the abated years, not the headline forgone revenue. Loudoun County is the clean illustration: data centers occupy roughly 4 percent of commercial parcels yet supply about 38 percent of general-fund revenue and

more than \$100 million in new revenue per year, enough to cut the real-property rate every year from \$1.145 in 2016 to \$0.805 in 2025 (Section 8). Where abatements are short and the tail of full-rate payments is long—and where the facility demands almost no offsetting services—the NPV can be positive even if the first-year optics look costly. This is the strongest case for a *sales-tax* exemption on business machinery, which Walczak (2025) argues is simply the sales tax operating as intended rather than a discriminatory carve-out. The distinction the paper insists on is between this neutral, non-discriminatory treatment of capital inputs and open-ended, firm-specific grants.

### **9.5 Comparison with manufacturing megadeals and reconciliation with audits**

Set against conventional manufacturing megadeals, data-center subsidies are simultaneously cheaper per dollar of capital and dearer per job, precisely because the capital-to-labor ratio is extreme. The race-to-the-bottom literature frames why this matters nationally: inter-state subsidy competition transferred roughly \$40 billion to firms for about \$13 billion in gains and would lower welfare less than 5 percent if banned outright (Slattery 2025), and state investment incentives largely relocate activity rather than create it (Chirinko and Wilson 2008). Two government audits corroborate the ledger from the cost side. Virginia forgoes on the order of \$1.6 billion in state sales-and-use tax annually (about \$1.9 billion with local) and recovers an estimated 48 cents per subsidy dollar, at the upper end of the 30-to-48-cent range that states report across the country (Tarczynska and LeRoy 2025; JLARC 2024). Georgia's audited evaluation, applying a 30 percent but-for attribution, found the high-tech data-center exemption net fiscally negative throughout—about \$474 million forgone in FY2025, rising toward \$867 million by 2030 (Carl Vinson Institute of Government 2025). That two independent state bodies reach the same sign as the MVPF framework, from different data, is the section's strongest external validation.

## 10. Counterarguments and Externalities

A favorable verdict on data-center *investment* is not a favorable verdict on every consequence of that investment. Three objections deserve to be taken at full strength rather than dismissed: that large computing loads shift electricity costs onto residential ratepayers, that data centers consume scarce water, and that they create too few permanent jobs to matter. Each contains something true. None, on the evidence assembled here, supports prohibition; each instead points toward better pricing and better incentive design—the same conclusion the cost-benefit ledger reaches from the fiscal side.

**Electricity and ratepayer cost-shifting.** The strongest version of the energy objection is not that data centers use a lot of power—that they do is uncontested, and a private input cost the firm itself bears is not an externality. The objection is distributional: that the fixed costs of generation, transmission, and interconnection induced by a new large load can be socialized across the rate base, so that households and small businesses pay for infrastructure built to serve a hyperscaler. Martin and Peskoe (2025), reviewing roughly fifty utility regulatory proceedings, document precisely this mechanism—discounted hookup terms, special contracts, and federal-transmission-versus-state-retail cost disconnects that let large loads underpay their marginal cost of service. The forward-looking magnitudes are nontrivial: JLARC (2024) projects that, absent rate reform, the typical Dominion residential bill in Virginia could rise by roughly \$14 to \$37 per month in constant dollars by 2040 on account of data-center load growth. That is a real number and the paper does not minimize it.

One qualification keeps the objection in proportion. The same JLARC (2024) analysis found *no historical cost-shift to date*: the projected harm is a contingent forecast about future rate design, not a measured transfer that has already occurred. Honest framing requires stating both halves—a real prospective risk and the absence, so far, of a realized one. New load is not mechanically a tax on incumbents; whether it becomes one depends on who is

assigned the incremental cost, which is a policy choice. The implication is therefore not a moratorium but **cost-causation pricing**—large-load tariffs, minimum-demand and ramp commitments, and direct assignment of interconnection costs to the party that triggers them—so that the firm enjoying the private benefit of cheap, reliable power also bears its private cost. The corrective lies in the rate structure, not in a prohibition on the load.

**Water.** The water objection is genuine but smaller and more localized than headline figures imply. LBNL (2024) estimates U.S. data centers consumed on the order of 17 billion gallons of water directly for cooling in 2023, plus roughly 211 billion gallons indirectly through thermoelectric power generation. The indirect figure is the larger one, and it is therefore in substantial part a *function of the generation mix*: a facility drawing from a grid with more renewables or closed-cycle cooling embeds less water than one drawing from open-loop thermoelectric plants. The externality is real where it bites—water-stressed regions such as Arizona and parts of Iowa, where the marginal gallon has a high shadow price—and negligible where water is abundant. This is an argument for **siting and pricing water at its scarcity value**, including reuse mandates and air- or recycled-water cooling in arid basins, not for a national restriction calibrated to the most stressed county. Where water is correctly priced, the firm internalizes the constraint.

**Too few jobs.** The most common popular objection—that a campus costing billions employs only a hundred-odd people—is true on its own terms and beside the point as stated. A 100-megawatt hyperscale facility runs on roughly 100–200 permanent staff, and the construction boom that precedes operation is by design transitory. If the case for data centers rested on a permanent-payroll multiplier in the Moretti (2010) sense, it would be weak: 150 jobs times a generous multiplier is still small. But the on-site headcount and the county-level effect are not the same object, and the apparent contradiction between the few-jobs critique and the favorable employment evidence dissolves once that is seen. Bahar and Wright (2026) find that a county’s first large data center raises *total private* employment by 4–5 percent and wages by 3–4 percent over five to

six years—but only in hyperscale clusters, with information-sector employment up some 23 percent in counties with four or more facilities and no significant effect at isolated single sites. The mechanism is agglomeration, the same force Greenstone, Hornbeck, and Moretti (2010) identify in incumbent-plant productivity, not direct hiring. The honest synthesis, then, is conditional and unembellished: data centers are a poor jobs program but, in clusters, a real capital-and-tax-base play with measurable spillovers. That conditionality is a feature of the argument, not an embarrassment to it—it tells policymakers exactly which projects (clustered, hyperscale) plausibly generate the spillovers that could justify any public role, and which (isolated colocation) do not. None of this rescues the subsidy; it sharpens where the burden of proof for one should fall.

## 11. Policy Implications

The two-ledger framework that organizes this paper yields a correspondingly two-part policy posture: welcome the investment, and reform the means used to attract it. Nothing in the evidence assembled here counsels against data-center development as such. The Loudoun County case (Section 8) and the broader employment evidence (Section 7) document a class of capital that deepens the local stock, broadens the assessed-value base, and imposes almost no marginal demand on schools, roads, or social services—an unusually clean instance of a tax base that pays for public goods it scarcely consumes. The case for reform attaches not to the plants but to the firm-specific carve-outs and rate structures layered on top of them. Four recommendations follow, each tied to a finding in the foregoing analysis.

**Prefer neutral, low taxation of business capital to firm-specific carve-outs.** The cleanest fiscal treatment of data centers is the one that requires no targeting at all. As Walczak (2025) argues, exempting machinery and equipment from the *sales* tax is the sales tax operating as intended—a destination tax on final consumption should not cascade onto business inputs—

and a state that broadly declines to tax capital purchases is not subsidizing anyone. That is categorically distinct from a discretionary, firm-conditioned property-tax abatement or a negotiated grant, which singles out one taxpayer for relief denied its neighbors. The incidence evidence sharpens the distinction: because a large share of any firm-specific tax break is captured by firm owners rather than local workers or landowners (Suarez Serrato and Zidar 2016), and because the marginal recipient is often an out-of-region shareholder, a targeted carve-out transfers local revenue to capital that, by the but-for logic of Section 9, would frequently have located there regardless. A uniform, low rate on business property achieves the locational competitiveness that boosters seek without the inframarginal leakage. The policy implication is to compete on the *general* tax environment, not on the *particular* deal.

**Replace open-ended exemptions with performance-tied, sunset-dated, clawback-backed incentives.** Where a jurisdiction insists on offering a discretionary incentive, its design should reflect the cost-benefit arithmetic of Section 9. The dominant data-center instrument—a sales-and-use-tax exemption running thirty or forty years and conditioned only on a one-time capital or job threshold—is precisely the open-ended form that the audited evidence finds value-destroying: Georgia’s evaluation reported a net-negative fiscal impact throughout the program horizon under a thirty-percent but-for attribution (Carl Vinson Institute of Government 2025), and states recover only an estimated thirty to forty-eight cents per subsidy dollar (Tarczynska and LeRoy 2025). The remedy is structural. Tie disbursement to verified, *induced* outcomes rather than promised ones; impose a hard sunset so the award does not outlive its rationale; and back every commitment with a clawback triggered by shortfalls in employment or investment. Bartik (2018b, 2019) shows that the cost per genuinely induced job—near \$200,000 under a central but-for rate of roughly twelve percent (Bartik 2018a)—is many times the headline cost per *promised* job, and that customized services cost a fraction as much per induced job as cash. A per-job cap of the kind Good Jobs First proposes—operationalized in its 2025 update as a \$50,000-per-job ceiling (Tarczynska and LeRoy 2025)—would mechanically prevent the \$1.4-to-\$2-million-per-job

outcomes documented for marquee data-center deals (Tarczynska and LeRoy 2016).

**Adopt cost-causation, large-load electricity tariffs and require disclosure of special contracts.** The ratepayer externality identified in Section 10 is a rate-design failure, not an argument against the load itself. Martin and Peskoe (2025) document how discounted hookup terms and confidential special contracts can shift grid costs onto residential customers, and JLARC (2024) projects typical Dominion residential bills rising \$14 to \$37 per month by 2040 absent reform—while finding no historical cost-shift yet, which is the fair framing. The corrective is to make large new loads bear the capacity and transmission costs they cause through dedicated large-load tariffs, minimum-demand and exit-fee provisions, and mandatory public disclosure of special-contract terms so regulators can verify that other classes are held harmless. This fixes the distributional problem without forgoing the investment.

**Mandate GASB-77-style subsidy disclosure as the precondition for any evaluation.** None of the above is enforceable without transparency. The recovery, but-for, and per-job figures this paper relies on exist only because some states publish tax-expenditure accounts; the analytical difficulty throughout has been that subsidy magnitudes are frequently undisclosed at the local level (Section 5). Routine, itemized reporting of every abatement and exemption—by recipient, program, and forgone revenue—is the minimal infrastructure that lets citizens and analysts run the cost-per-induced-job and break-even calculations of Section 9 *before* a forty-year commitment is signed rather than after.

*[Original estimates forthcoming from the county panel will quantify how much of the realized local tax-base and employment gain is attributable to the marginal facility versus inframarginal capital; the expected pattern, based on Bahar and Wright (2026) and the but-for literature (Bartik 2018a), is that the*

*favorable verdict holds for the investment and the unfavorable verdict for the blanket subsidy, conditional on facility type and clustering.]*

## **12. Conclusion**

This paper has argued for separating two questions that public debate over data centers routinely collapses into one. The first is whether the *investment* is good for the county that hosts it; the second is whether the *subsidy* used to attract it is good for the public that pays for it. The evidence points in opposite directions on the two, and conflating them produces both the booster's overclaim and the critic's overcorrection.

On the investment, the weight of the evidence is favorable. Data centers deliver large capital deepening and a property- and sales-tax base that demands almost no public services—the Loudoun County pattern, in which facilities on a small share of commercial parcels fund a disproportionate share of the general fund, is close to a textbook public-finance result (JLARC 2024). The best causal employment evidence finds that a county's first large facility raises total private employment on the order of 4–5 percent and wages 3–4 percent over five to six years, with the gains concentrated in hyperscale clusters rather than isolated colocation sites (Bahar and Wright 2026). That is consistent with the agglomeration-spillover tradition (Greenstone, Hornbeck, and Moretti 2010) and with modest local multipliers (Moretti 2010) operating on a small permanent base. Welcoming the plants, and taxing their capital neutrally rather than pentially (Walczak 2025), is the defensible position.

On the subsidy, the evidence does not support open-ended, firm-specific carve-outs. Because most subsidy dollars are inframarginal—paid for investment that would have occurred anyway—the welfare-relevant cost per *induced* job is roughly an order of magnitude larger than the cost-per-promised-job figure boosters cite (Bartik 2018a, 2018b), and the resulting marginal value of public funds is low (Hendren and Sprung-Keyser 2020). Competitive bidding compounds the loss by transferring surplus to firm owners, often out of region,

with little net national gain (Slattery 2025; Suarez Serrato and Zidar 2016; Chirinko and Wilson 2008).

These conclusions carry real limits, and one in particular should be stated plainly: they rest on the external causal literature, because the paper's own county-panel and synthetic-control estimates remain forthcoming and are flagged as such throughout. The favorable verdict is conditional on facility type and clustering, not general; effects are heterogeneous and may partly reflect activity reshuffled across jurisdictions rather than created nationally (Kline and Moretti 2014b). External validity beyond the studied counties is uncertain, NAICS 518210 undercounts captive hyperscale capacity, and county-level suppression constrains the usable sample. [*Original estimates forthcoming from the county panel; expected sign and magnitude follow Bahar and Wright (2026) and the spillover and multiplier literatures, with formal sensitivity to parallel-trends violations per Rambachan and Roth (2023).*] The free-market case here is not that subsidies are harmless, but that the investment can stand on its own—so the policy that maximizes welfare is to welcome the plants and retire the abatements.

## References

- Alberto Abadie, Alexis Diamond, and Jens Hainmueller. 2010. "Synthetic Control Methods for Comparative Case Studies: Estimating the Effect of California's Tobacco Control Program." *Journal of the American Statistical Association*, 105(490), 493-505 (DOI 10.1198/jasa.2009.ap08746).
- Arman Shehabi, Sarah J. Smith, Alex Hubbard, et al. (Lawrence Berkeley National Laboratory). 2024. "2024 United States Data Center Energy Usage Report." *Lawrence Berkeley National Laboratory, for the U.S. Department of Energy (LBNL-2024)*.
- Joint Legislative Audit and Review Commission (JLARC), Commonwealth of Virginia. 2024. "Data Centers in Virginia (Report 598)." *Virginia General Assembly (JLARC), December 2024*.
- Dany Bahar and Greg C. Wright. 2026. "New Evidence on Data Center Employment Effects." *Brookings Institution (commentary on synthetic-control working paper, ~770 facilities, 93 treated counties, 2003-2024)*.

- Timothy J. Bartik. 1991. "Who Benefits from State and Local Economic Development Policies?." *W.E. Upjohn Institute for Employment Research*.
- Timothy J. Bartik. 2017. "A New Panel Database on Business Incentives for Economic Development Offered by State and Local Governments in the United States (PDIT)." *W.E. Upjohn Institute, report prepared for the Pew Charitable Trusts*.
- Timothy J. Bartik. 2018. "'But For' Percentages for Economic Development Incentives: What Percentage Estimates Are Plausible Based on the Research Literature?." *W.E. Upjohn Institute, Working Paper 18-289*.
- Timothy J. Bartik. 2018. "The Bartik Benefit-Cost Model of Business Incentives (User's Guide; Introduction)." *W.E. Upjohn Institute, Reports 287 and 288*.
- Timothy J. Bartik. 2019. "Making Sense of Incentives: Taming Business Incentives to Promote Prosperity." *W.E. Upjohn Institute for Employment Research (DOI 10.17848/9780880996693)*.
- Brantly Callaway and Pedro H. C. Sant'Anna. 2021. "Difference-in-Differences with Multiple Time Periods." *Journal of Econometrics*, 225(2), 200-230.
- Carl Vinson Institute of Government, University of Georgia (for the Georgia Dept. of Audits and Accounts). 2025. "Tax Incentive Evaluation: High-Tech Data Center Sales and Use Tax Exemption." *Georgia Department of Audits and Accounts*.
- Clement de Chaisemartin and Xavier D'Haultfoeuille. 2020. "Two-Way Fixed Effects Estimators with Heterogeneous Treatment Effects." *American Economic Review*, 110(9), 2964-2996 (NBER WP 25904).
- Clement de Chaisemartin and Xavier D'Haultfoeuille. 2023. "Two-Way Fixed Effects and Differences-in-Differences with Heterogeneous Treatment Effects: A Survey." *The Econometrics Journal*, 26(3), C1-C30 (NBER WP 29691).
- Clement de Chaisemartin and Xavier D'Haultfoeuille. 2024. "Difference-in-Differences Estimators of Intertemporal Treatment Effects." *The Review of Economics and Statistics (DOI 10.1162/rest\_a\_01414; NBER WP 29873)*.
- Robert S. Chirinko and Daniel J. Wilson. 2008. "State Investment Tax Incentives: A Zero-Sum Game?." *Journal of Public Economics*, 92(12), 2362-2384 (FRBSF WP 2006-47).
- Dmitry Arkhangelsky, Susan Athey, David A. Hirshberg, Guido W. Imbens, and Stefan Wager. 2021. "Synthetic Difference-in-Differences." *American Economic Review*, 111(12), 4088-4118 (DOI 10.1257/aer.20190159; NBER WP 25532).
- Jared Walczak (Tax Foundation). 2025. "The Taxation of Data Centers." *Tax Foundation research report*.
- Edward L. Glaeser and Joshua D. Gottlieb. 2008. "The Economics of Place-Making Policies." *Brookings Papers on Economic Activity*, 2008(1), 155-239 (NBER WP 14373).

- Edward L. Glaeser and Joshua D. Gottlieb. 2009. "The Wealth of Cities: Agglomeration Economies and Spatial Equilibrium in the United States." *Journal of Economic Literature*, 47(4), 983-1028 (NBER WP 14806).
- Andrew Goodman-Bacon. 2021. "Difference-in-Differences with Variation in Treatment Timing." *Journal of Econometrics*, 225(2), 254-277 (NBER WP 25018).
- Nathaniel Hendren and Ben Sprung-Keyser. 2020. "A Unified Welfare Analysis of Government Policies." *Quarterly Journal of Economics*, 135(3), 1209-1318 (NBER WP 26144).
- Jesse Noffsinger, Mark Patel, and Pankaj Sachdeva (McKinsey & Company). 2025. "The Cost of Compute: A \$7 Trillion Race to Scale Data Centers." *McKinsey & Company*.
- Kirill Borusyak, Xavier Jaravel, and Jann Spiess. 2024. "Revisiting Event-Study Designs: Robust and Efficient Estimation." *Review of Economic Studies*, 91(6), 3253-3285 (DOI 10.1093/restud/rdae007).
- Patrick Kline and Enrico Moretti. 2014. "People, Places, and Public Policy: Some Simple Welfare Economics of Local Economic Development Programs." *Annual Review of Economics*, 6, 629-662 (NBER WP 19659).
- Patrick Kline and Enrico Moretti. 2014. "Local Economic Development, Agglomeration Economies, and the Big Push: 100 Years of Evidence from the Tennessee Valley Authority." *Quarterly Journal of Economics*, 129(1), 275-331 (NBER WP 19293).
- Eliza Martin and Ari Peskoe. 2025. "Extracting Profits from the Public: How Utility Ratepayers Are Paying for Big Tech's Power." *Harvard Law School, Environmental and Energy Law Program (Harvard Electricity Law Initiative, March 2025)*.
- Matias Busso, Jesse Gregory, and Patrick Kline. 2013. "Assessing the Incidence and Efficiency of a Prominent Place Based Policy." *American Economic Review*, 103(2), 897-947 (NBER WP 16096).
- Michael Greenstone, Richard Hornbeck, and Enrico Moretti. 2010. "Identifying Agglomeration Spillovers: Evidence from Winners and Losers of Large Plant Openings." *Journal of Political Economy*, 118(3), 536-598 (NBER WP 13833).
- Enrico Moretti. 2010. "Local Multipliers." *American Economic Review: Papers & Proceedings*, 100(2), 373-377.
- Enrico Moretti. 2011. "Local Labor Markets." *Handbook of Labor Economics*, Vol. 4B, Ch. 14, pp. 1237-1313 (Elsevier; NBER WP 15947).
- David Neumark and Helen Simpson. 2015. "Place-Based Policies." *Handbook of Regional and Urban Economics*, Vol. 5B, Ch. 18, pp. 1197-1287 (Elsevier; NBER WP 20049).
- Ashesh Rambachan and Jonathan Roth. 2023. "A More Credible Approach to Parallel Trends." *Review of Economic Studies*, 90(5), 2555-2591 (DOI 10.1093/restud/rdado18).

- Jonathan Roth. 2022. "Pretest with Caution: Event-Study Estimates after Testing for Parallel Trends." *American Economic Review: Insights*, 4(3), 305-322 (DOI 10.1257/aeri.20210236).
- Juan Carlos Suarez Serrato and Owen Zidar. 2016. "Who Benefits from State Corporate Tax Cuts? A Local Labor Markets Approach with Heterogeneous Firms." *American Economic Review*, 106(9), 2582-2624 (NBER WP 20289).
- Cailin Slattery and Owen Zidar. 2020. "Evaluating State and Local Business Incentives." *Journal of Economic Perspectives*, 34(2), 90-118 (NBER WP 26603).
- Cailin Slattery. 2025. "Bidding for Firms: Subsidy Competition in the United States." *Journal of Political Economy*, 133(8), 2563-2614 (DOI 10.1086/735509; SSRN 3250356).
- Liyang Sun and Sarah Abraham. 2021. "Estimating Dynamic Treatment Effects in Event Studies with Heterogeneous Treatment Effects." *Journal of Econometrics*, 225(2), 175-199.
- Kasia Tarczynska and Greg LeRoy (Good Jobs First). 2016. "Money Lost to the Cloud: How Data Centers Benefit from State and Local Government Subsidies." *Good Jobs First (think-tank report)*.
- Kasia Tarczynska and Greg LeRoy (Good Jobs First). 2025. "Cloudy with a Loss of Spending Control: How Data Centers Are Endangering State Budgets." *Good Jobs First (think-tank report, April 2025)*.
- John Douglas Wilson. 1999. "Theories of Tax Competition." *National Tax Journal*, 52(2), 269-304 (DOI 10.1086/NTJ41789394).

---

*This is a draft working paper prepared for an AIER summer research project. The views expressed are the author's and do not necessarily reflect those of the American Institute for Economic Research. Original county-panel estimates are forthcoming and are flagged as such in the text; quantitative claims otherwise derive from the cited literature.*