Summary Memo

Do Hyperscale Data Centers Raise US Electricity Prices?

Lisa Thalheimer, Kevin Butler & Francesca Dominici

Prepared by Lisa Thalheimer, DPhil, Research Associate, Harvard T.H. Chan School of Public Health

Introduction

In brief: AI-driven data center growth is sharply raising US electricity demand, and this project investigates whether it is already raising prices, using five related research questions covering price, dose-response, developer type, spatial spillovers, and local economic benefits.

Hyperscale data centers built to train and run artificial intelligence (AI) systems are driving one of the fastest sustained increases in US electricity demand in decades. The International Energy Agency (IEA) projects global data center electricity consumption to roughly double by 2030, with AI-optimized facilities driving most of that growth and data centers accounting for nearly half of US electricity demand growth over the same period (IEA, Energy and AI, 2025). Whether this demand growth is already raising the prices households and businesses pay is contested. Recent causal analyses reach opposite conclusions depending on geographic scope and time horizon, with one finding a modest local price increase concentrated in industrial rates (Rana et al., 2026) and another finding a modest decline in the national average retail rate (Watten et al., 2026). Beyond price, a parallel literature has begun documenting environmental and health concerns tied to facility siting, including air pollution from fossil-fuel generation and backup diesel generators, water use for cooling, and noise. Virginia's dense data center corridor has served as an early case study of these concerns (Gour, 2026), and facility-level modeling shows that how a campus sources its power materially shifts who bears the resulting air-quality burden (Gujral & Dominici, 2026).

Here we ask four research questions. (1) Do hyperscale data centers raise the retail electricity prices paid by industrial and residential customers of the utilities that host them? (2) Does the size of a utility's data center exposure, beyond its presence, predict a larger price effect? (3) Does the type of developer, large hyperscalers building their own campuses versus other developers, change the size of any effect? (4) Does a utility's price respond to data center capacity sited in neighboring utility territories, not just its own? A companion analysis asks a fifth, related question at the county level. (5) Do host counties show any measurable local economic benefit, such as higher employment or income, that might offset these costs?

Data Overview

In brief: we combine data center location and capacity records with government data on electricity prices, jobs, and income to build a dataset covering US utilities and counties from 2015 to 2025.

We use two main data sources, linked through each utility’s federal ID number. The first is CleanView, a private dataset that tracks hyperscale data centers across the country: their location, size, owner, and the year each one opened. We include only campuses of at

least 50 megawatts (1,373 campuses meet this bar; 443 are already operating). The second is the government's EIA-861 "Sales to Ultimate Customers" dataset, which reports real electricity prices, sales, and customer counts by utility, state, and year from 2021 to 2025 (11,606 rows across 1,602 utilities). This replaces an earlier, unreliable sample we started with.

We also draw on several supporting sources: utility service-area maps (from a federal infrastructure database) to study how a utility's price relates to its neighbors; U.S. county boundaries to link each data center to the county it sits in; and, for a companion analysis of local economic benefits, government employment data (BLS) and income data (BEA) covering counties from 2015 to 2024, a longer stretch of time than the price data allows. We also built two measures to test whether other factors could explain our results: how much "retail choice" exists in a utility's electricity market, and wholesale electricity prices in the mid-Atlantic/Midwest grid region (PJM). Finally, we checked whether a public tracker of data centers' power sources (grid-connected vs. self-generated) could support further analysis, but found its data too sparse and unreliable to use (see Discussion).

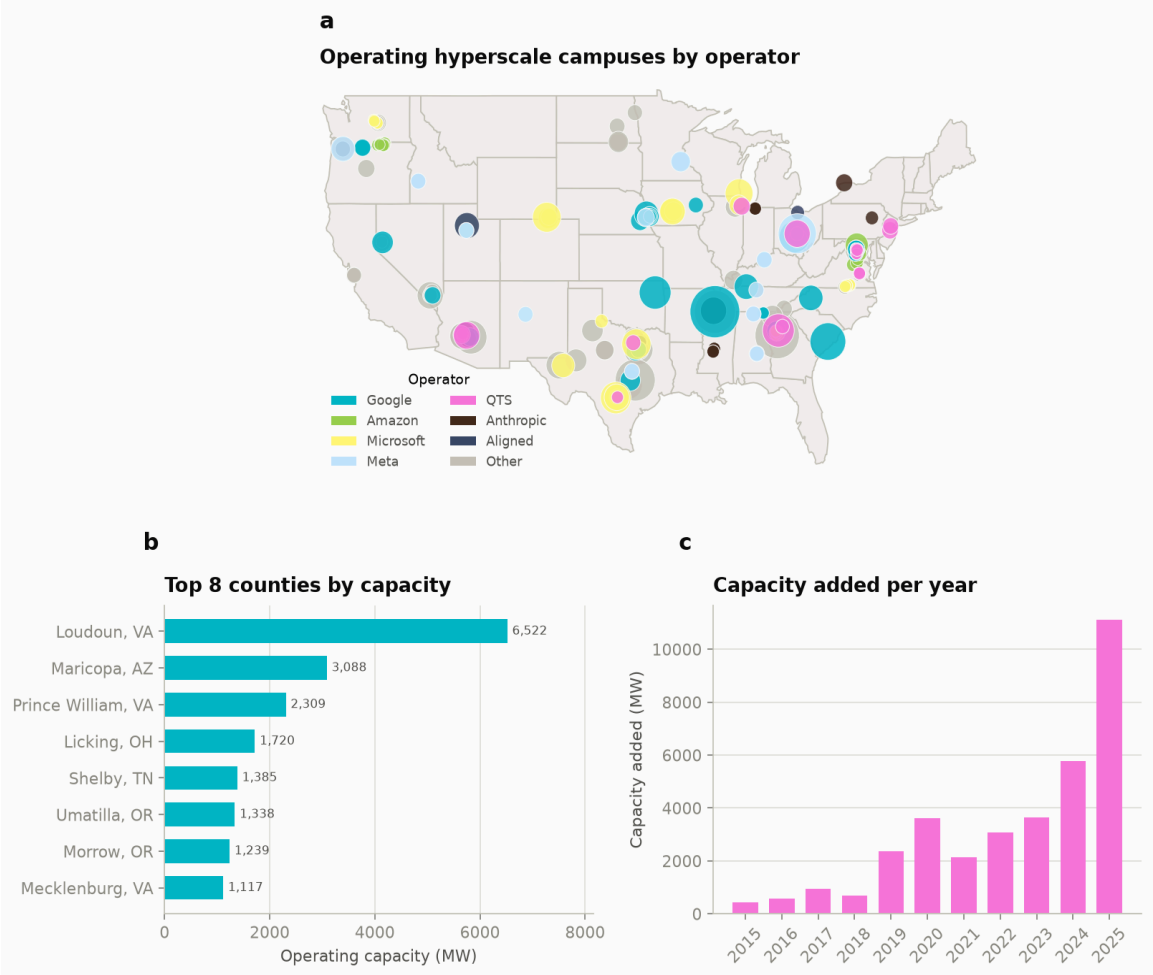


Figure 1. Where hyperscale data centers are located, who operates them, and how fast capacity is being added (CleanView, version 10/07/2026, Operating campuses ≥ 50 MW).

One more caveat about Figure 1 is worth flagging plainly: how concentrated data-center capacity looks depends heavily on the size cutoff used. Figure 1 counts every campus of at least 50 megawatts, and by that count Virginia leads every other state by a wide margin. But if we only count true mega-campuses (at least 100 megawatts), Texas overtakes Virginia for the top spot (4,711 vs. 3,171 megawatts), and the ranking below

them reshuffles too. That's because Virginia's lead is built mostly from many mid-sized campuses in the 50–99 megawatt range, not a handful of giant ones: at the stricter cutoff, the share of Virginia's capacity held by the seven largest operators drops sharply (from 61% to 20%), and two companies prominent at the lower cutoff, Amazon and Anthropic, drop out of the national top 7 operators entirely. This doesn't change any of our price results, which depend on whether a utility hosts any qualifying campus, not on a state-level capacity count, but it does mean "which state has the most data centers" is a modeling choice, not a settled fact.

Models

In brief: we treat the opening of each data center as an event and compare prices and jobs before and after, using a statistical method built for comparing places that get "treated" at different times, cross-checked with a simpler backup method and several additional tests.

Our main approach compares a utility or county's prices before and after its first hyperscale campus opens, treating that opening as the "treatment." Because campuses open in different years, we use a statistical method (Callaway & Sant'Anna, 2021, called "CS" below) built specifically for this kind of staggered timing, chosen because it is designed to give trustworthy results even with very few treated cases, though with only 12 treated utilities in our main price analysis, any result here should be read cautiously. We cross-check it against a simpler, more transparent method (Cengiz et al., 2019), and also report what a naive before/after comparison would show, mainly to illustrate the bias that method introduces when treatment timing is staggered (Goodman-Bacon, 2021).

Beyond the main comparison, we ran several extensions: testing whether bigger data centers have bigger effects, not just whether one exists nearby; comparing hyperscale companies that build their own data centers against other developers; testing whether a utility's price is affected by its neighbors' data center capacity; and applying the same before/after comparison to county-level jobs and income, including a later addition testing total county employment against outside research. We also ran two checks to see whether our results could be explained by other factors: adding a measure of retail electricity competition, and narrowing the sample to the mid-Atlantic/Midwest grid region where we could also control for wholesale electricity prices.

Results

In brief: industrial prices rise after a data center opens, but not by a reliable amount; an apparent drop in residential prices turns out to be an artifact of a flawed comparison; the clearest new finding is that a utility's price rises when its neighbors host more data center capacity; and local job and income effects are either undetectably small or unconfirmed.

Industrial electricity prices rise after a data center opens in every version of our analysis (by $13 to $51 per megawatt-hour, depending on how long after opening we look), but the effect is never statistically reliable. With only 12 treated utilities in our cleanest comparison, we simply don't have enough cases to tell whether the true effect is small, zero, or genuinely large.

Our main method also finds a statistically significant drop in residential prices after a data center opens. But this comparison requires that treated and untreated utilities were on similar price trends beforehand, and we found a significant difference between them three years before any data center opened. That means the "drop" likely reflects a flawed comparison, not a real effect of the data center, and should not be read as evidence that data centers lower residential prices.

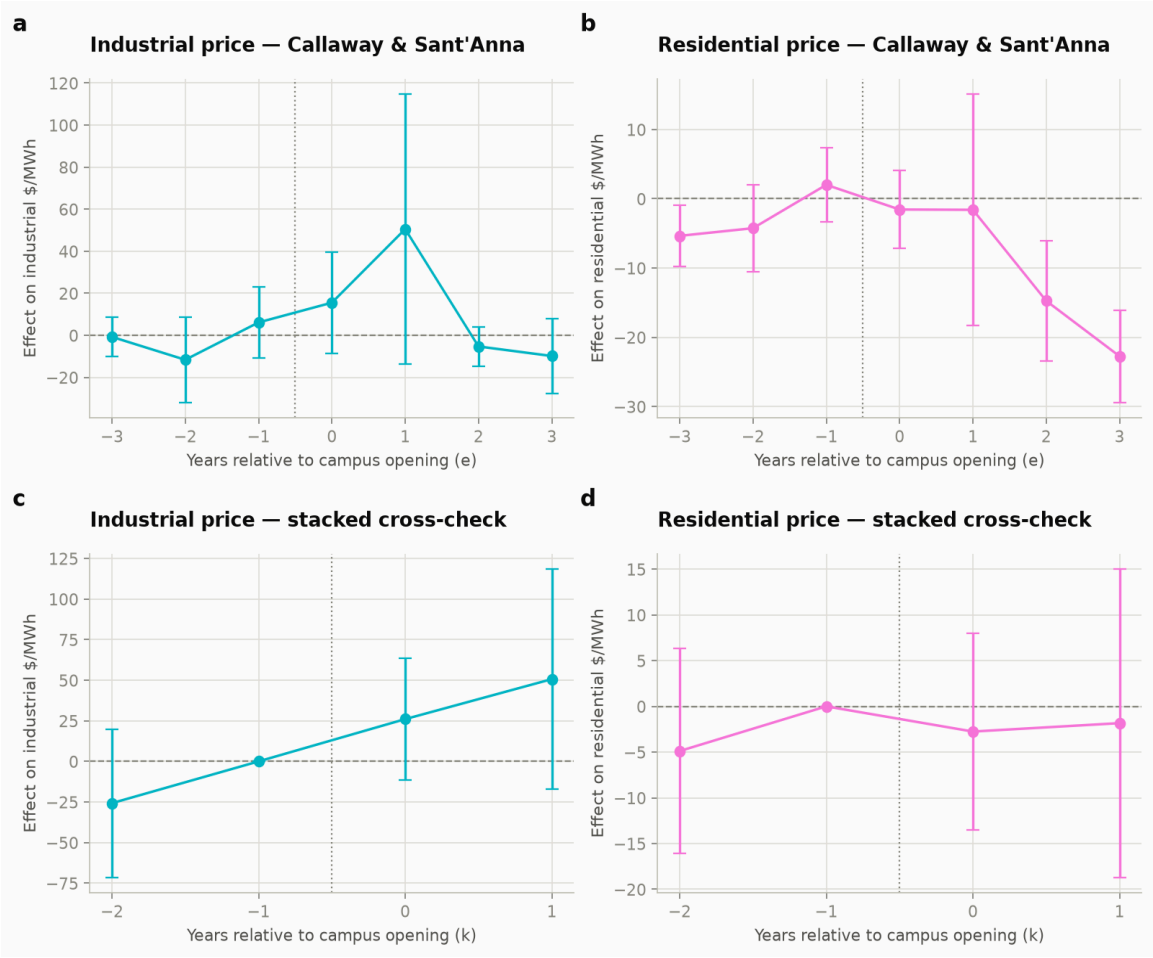


Figure 2. Industrial and residential price effects around a campus opening, primary (CS) estimator and stacked cross-check. Panel (b) shows the pre-period point that invalidates the apparent residential decline.

The clearest new result: a utility's own residential price is higher when its neighboring utilities host more data center capacity (statistically significant, in a much larger sample of 1,199 utilities). This is only a correlation, not proof that one causes the other (something else about a region could be driving both), but it is the one place our data disagrees with prior published research, which found no such neighbor effect.

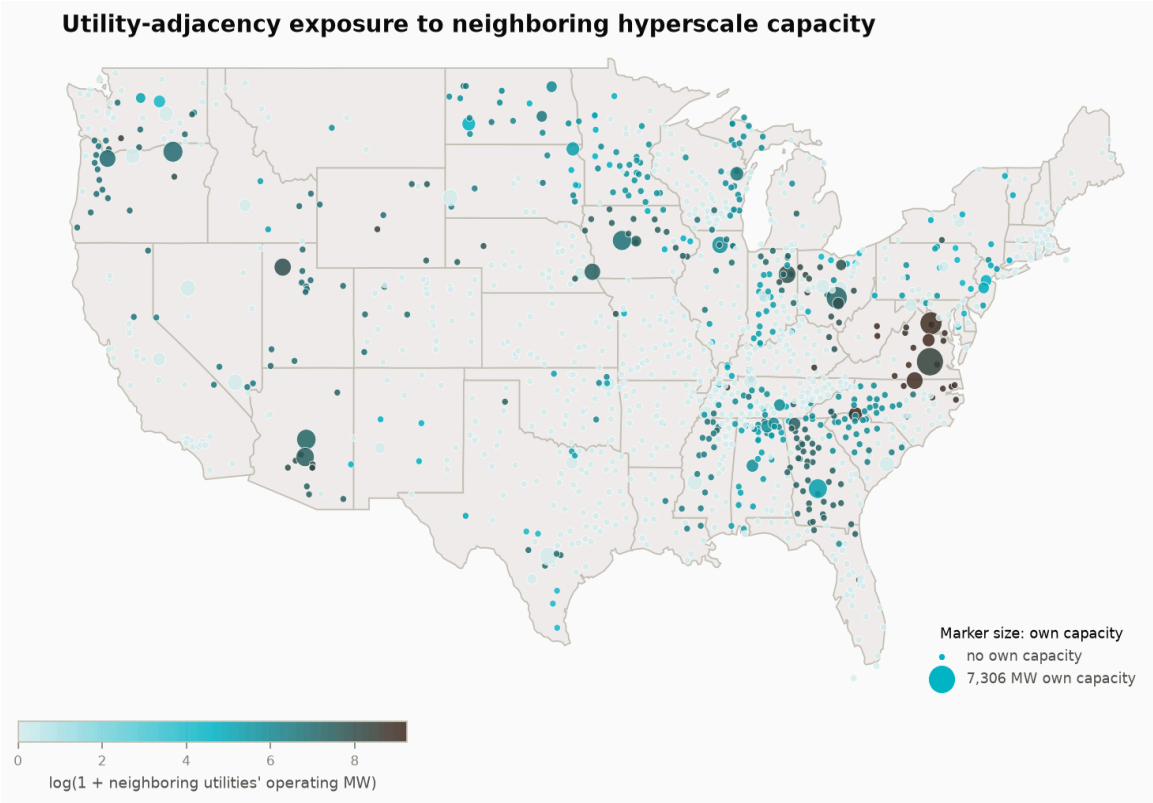


Figure 3. Utility-adjacency exposure to neighboring hyperscale capacity: the basis for the one result where this analysis differs from prior published research.

On the benefit side, we found no statistically detectable effect of a new data center on local data center-industry jobs or per-capita income, in a sample of 28–35 treated counties. That doesn't prove there's no effect. It just means our data can't rule much in or out. We also tested a broader measure, total county employment, because outside research (Yue & Zeng) found a national 3.5% employment bump from data center openings. Our data showed a similar rise after openings, but a separate check also found a suspicious difference before openings, so this result is intriguing but unconfirmed and not reliable.[1]

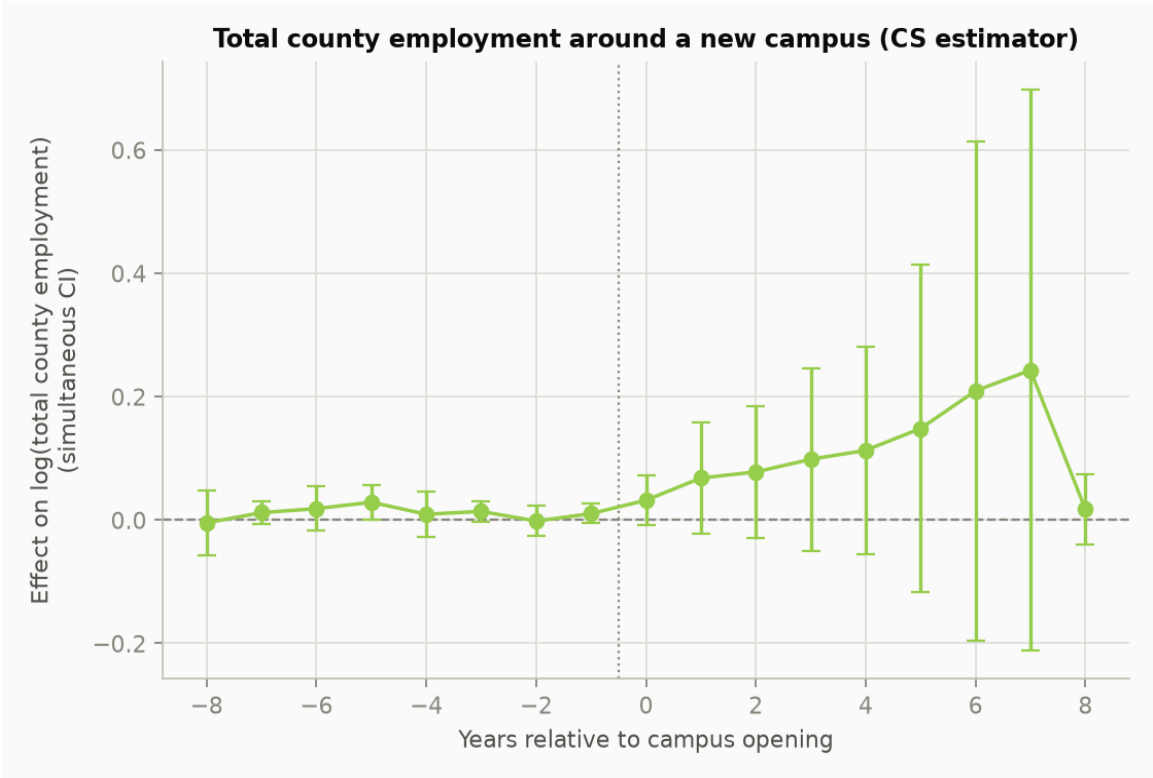


Figure 4. Total county employment around a campus opening (CS estimator). A rising post-period pattern, but widening confidence intervals that eventually cross zero, suggestive but not confirmed.

Two further checks turned up what looked like results but didn't hold up: splitting the data by type of developer, and checking grid-reliability outcomes. Both showed patterns best explained by having very few cases to compare, not a real effect, so we report them without treating them as findings. A check for electricity-market competition (retail choice) didn't change our main results, but is itself worth noting: it predicted lower prices when measured one way, and higher prices when measured another way, a sign that this particular factor is not well understood, not evidence about competition's real effect on price.

Our subsample covering the mid-Atlantic/Midwest grid region (PJM) found a much larger, statistically significant price effect, roughly six times the national estimate. But it rests on only 7 utilities, 5 of which were already hosting a data center before our data begins, so it likely isn't generalizable.

One further limitation worth flagging plainly: we could not test whether it matters if a data center draws its power from the grid or generates its own ("behind-the-meter"), because no dataset, including our own, reliably tracks this. A public tracker we tested miscoded a well-known self-powered data center as grid-connected. This means any self-powered campuses in our data are being counted the same as grid-connected ones, which likely weakens our results rather than invalidating them.

Discussion and Lessons Learnt

In brief: our findings line up with, but don't prove, recent regulatory concerns that data centers shift costs onto other electricity customers. The biggest lessons from this project are about having too little data to be certain, and about a basic mismatch between who pays for data centers and who benefits from them.

Two recent studies on this topic reach opposite conclusions, but the difference makes sense once you look at what each one actually measures: local, recent effects near data centers (our work, and a paper by Rana et al.) versus a broad national average over a full decade (a paper by Watten et al.). These are different questions, not contradictory answers.

Real-world regulatory activity happening at the same time as our analysis points the same direction as our one clear finding. Maryland's utility regulator has filed a complaint with federal regulators (FERC) alleging that PJM, the regional grid operator, spreads the cost of data center-driven grid upgrades across the whole region rather than charging the data centers themselves (an allegation, not yet decided). A separate advocacy-group report raises the same concern. Our own finding, that a utility's price rises when its neighbors host more data center capacity, is the one place our evidence points the same way as that regulatory concern, even though we can't prove the same mechanism is at work.

A related state-level study by Scioto Analysis (a policy think tank, not peer-reviewed) modeled the costs and benefits of data centers in Ohio. It found that between 2027 and 2034, air pollution would cost Ohioans more (about $41 billion) than higher electricity prices (about $16 billion). So electricity price, while real, isn't actually the biggest cost in their model. Still, their estimated electricity cost alone was enough to push new data center construction to barely break even on a cost-benefit basis. Their numbers use a borrowed national price-growth estimate, not a causal estimate like ours, so the two studies complement rather than confirm each other. Their strongest recommendation, requiring new data centers to generate their own power instead of drawing from the grid, would cut that estimated electricity cost roughly fourfold; they note that roughly a fifth of Ohio's planned data center capacity is already headed that way.

A separate national-scale modeling exercise by Duke University's Nicholas Institute points the same way through a completely different method: rather than estimating realized prices, it simulates the entire US wholesale electricity market out to 2050 and finds that letting new data centers supply their own power ("bring-your-own-power") lowers average generation costs for everyone else still drawing from the shared grid, a pattern that holds up across a range of assumptions about natural gas prices and power-plant construction costs. It shares no data or method with either our study or Scioto's, so it isn't confirmation in a statistical sense — but a causal estimate (ours), a state-level cost accounting exercise (Scioto's), and a national market simulation (Duke's) all now point toward the same policy lever: moving new data-center load off the shared grid reduces the cost it imposes on everyone else.

That raises an important limitation of our own analysis: neither CleanView nor our own data tracks whether a campus draws power from the grid or generates its own. This means self-powered campuses in our data get counted exactly the same as grid-connected ones, which probably weakens (not invalidates) our price estimates. Separate research (Gujral & Dominici, a 2026 preprint) reinforces that this distinction matters for a different reason: it finds that a self-powered data center's air-pollution health harm concentrates locally, disproportionately on lower-income and minority communities, while a similar grid-connected campus's pollution harm spreads more broadly across distant communities. That's a different kind of harm than what we study, and doesn't validate our price findings, but it is strong independent evidence that how a data center gets its power is not a minor detail.

A co-author's review raised a useful challenge: instead of extending this comparison on its own terms, trace the chain from a utility's underlying costs to the price a household actually pays. Regulators approve generation, transmission, and distribution/operating costs for recovery, allocate them across customer classes, and set them into rate design

before they reach a retail bill; the question becomes whether data-center load moves price beyond what that chain already explains. Our data covers this chain unevenly: we already have wholesale generation prices for part of our sample (the PJM subsample above); transmission cost is not measured at all, despite being the specific mechanism in the Maryland complaint above; and distribution cost, operating expense, and the utility's rate base are not in our main data source at all — they live in a separate federal filing, FERC Form 1. We checked that filing's coverage specifically: our larger investor-owned utilities (Dominion, Commonwealth Edison, and similar) file it, but 6 of our 12 clean-cohort treated utilities — the rural electric cooperatives and the public utility district among them — are legally exempt from filing it at all, not just missing by chance. So this is a real, worthwhile extension, but one that can only ever cover about half our treated utilities, and it's a complement to our main comparison, not a replacement for it: our main design already accounts for unmeasured cost differences as long as treated and untreated utilities were on similar trends beforehand.

Key lessons from this project:

1. Using a statistical method actually designed for this kind of staggered comparison (CS), rather than a simpler, more common method, mattered: it caught a hidden pre-existing trend that the simpler method would have missed, turning an unclear result into a clear warning sign. Even so, with only 12 treated utilities, no method here can be considered fully validated.

2. The biggest obstacle throughout this project is having too few treated cases, especially on the price side (12 utilities). Extending our price data back to 2015 (we currently start in 2021) is the single most useful next step, and our longer county-level data already shows what that buys: not a clean “no effect,” but a real, if unconfirmed, signal in total employment that our shorter price data simply can’t see either way.

3. “We can’t detect an effect” and “there is no effect” are different claims, and we tried to keep them separate throughout, for example by not presenting an interesting-looking but unreliable result (the developer-type split) as a real finding.

4. We did not build a way to rule out “reverse causation” (the possibility that data centers choose to locate where power is already cheap, rather than making it cheap or expensive themselves), we flag this as an open gap, not something we worked around.

5. The costs of data centers (measured at the utility level, which can span a huge area) and the benefits (jobs, measured at the county level) are counted over different-sized areas entirely, and that mismatch matters on its own, regardless of how much data we have. We can even count it directly: of the 84 counties that host an operating data center, 7 (8%) are split between more than one electricity utility, including Loudoun County, Virginia, split between two utilities with 55 and 22 campuses respectively. This almost certainly understates the real mismatch, since it misses the more common case: a utility’s price effect reaching into counties that host no data center at all.

Policy Recommendations

In brief: based on these findings, we recommend better public disclosure of how data centers get their power, cost-allocation reform so nearby communities don’t subsidize distant data centers, conditions on tax breaks tied to real local benefits, and continued investment in longer, more detailed public data.

1. Require and standardize disclosure of how large data centers get their power (grid-connected vs. self-generated). No public or private dataset we could find, including our own, reliably tracks this, and the one tracker that attempts it misclassifies even well-known examples. This single gap affects both electricity-price analysis (self-powered campuses may not raise grid costs at all) and public health (self-generation concentrates air pollution locally rather than spreading it across the grid). Regulators overseeing large-load interconnections are well placed to require this disclosure.

2. Reform how the costs of data center-driven grid upgrades are allocated, rather than letting them spread automatically across a whole region. Our finding that a utility's price rises with its neighbors' data center capacity, together with an active regulatory complaint making a similar allegation, suggests costs may currently fall on communities that host no data center at all. Allocating costs closer to where the demand originates would be a more direct fix than broad price caps or blanket restrictions on data center construction.

3. Tie data center tax incentives to verified local benefits, with built-in review. Ohio's data center sales-tax exemption alone is estimated to cost the state roughly $1.6 billion a year in foregone revenue, while our own analysis finds no confirmed local job or income benefit yet. Incentives granted for up to 40 years with no re-evaluation are a mismatch for a technology and market this young; shorter terms with periodic reassessment, and incentives conditioned on measurable local hiring or investment, would better match the actual, still-uncertain, evidence of local benefit.

4. Encourage, rather than discourage, data centers that bring their own power, but pair that encouragement with local air-quality safeguards. Outside modeling (Scioto Analysis) finds that requiring new campuses to self-generate could substantially cut their electricity-price cost to the public. But separate research (Gujral & Dominici) finds that self-generated power concentrates air pollution locally, often on lower-income and minority communities. Policy that favors self-generation without requiring environmental review of the generation itself risks trading one problem (electricity prices) for another (local air quality).

5. Invest in better public data before legislating further. Nearly every uncertain or unreliable result in this analysis traces back to having too little data: electricity price records that only go back to 2021, no reliable way to know whether a data center is grid-connected or self-powered, and no routine way to know whether a region's data center capacity and its electricity prices are moving together over a longer period. Extending government electricity-price reporting further back in time, and requiring basic power-source disclosure, would let policymakers and researchers resolve these questions with actual evidence rather than the best guess currently available.

References

In brief: references cover the two primary comparison studies, the statistical methods used, regulatory and policy documents, and the underlying government and third-party data sources.

[1] Rana, V., Kaps, C. & Netessine, S. The Cost of Intelligence: Impacts of AI Data Centers on US Electricity Prices. Harvard Business School Working Paper (2026). papers.ssrn.com/sol3/papers.cfm?abstract_id=7580459

[2] Watten, A., Bistline, J. & Blanford, G. Have Data Centers Raised Your Electric Bill? Causal Evidence from the United States. arXiv:2606.19777 (2026). arxiv.org/abs/2606.19777

[3] Callaway, B. & Sant'Anna, P. H. C. Difference-in-Differences with multiple time periods. Journal of Econometrics 225, 200–230 (2021).

[4] Goodman-Bacon, A. Difference-in-differences with variation in treatment timing. Journal of Econometrics 225, 254–277 (2021).

[5] Cengiz, D., Dube, A., Lindner, A. & Zipperer, B. The effect of minimum wages on low-wage jobs. Quarterly Journal of Economics 134, 1405–1454 (2019).

[6] Energy Futures Group. Large Load Cost Allocation: Framework for States to Reduce Cross-Subsidization and Risk to Existing Customers. Commissioned by Earthjustice (2026). (Policy report commissioned by an advocacy organization, not peer-reviewed.) earthjustice.org/experts/shannon-fisk/new-report-examines-how-ai-data centers-are-increasing-electricity-costs

[7] Maryland Office of People's Counsel. Complaint against PJM Interconnection, L.L.C. regarding RTEP cost allocation for data center-driven transmission upgrades. Federal Energy Regulatory Commission, Docket No. EL26-63 (2026). (Regulatory complaint; allegations not yet adjudicated.) utilitydive.com/news/maryland-ratepayer-advocate-ferc-data center-complaint-transmission/823244

[8] Scioto Analysis. Data Center Policy in Ohio: A Cost-Benefit Analysis (2026). (Policy report, not peer-reviewed.)

[9] Ross, M. T. & Ewing, J. Controlling Impacts of Data Centers on Electricity Generation Costs: A Modeling Analysis. NI 26-24. Nicholas Institute for Energy, Environment & Sustainability, Duke University (2026). (Policy report, not peer-reviewed.) nicholasinstitute.duke.edu/publications/controlling-impacts-data-centers-electricity-generation-costs

[10] Gujral, H. & Dominici, F. Behind-the-Meter Versus Grid Electricity for AI Datacenters: Magnitude, Geography, and Equity of PM2.5-Attributable Mortality in Memphis. Preprints.org (2026). doi.org/10.20944/preprints202610.0819.v1

[11] FracTracker Alliance. US Data Centers Tracker. (Advocacy-nonprofit data tracker, not a peer-reviewed source.) fractracker.org

[12] Environmental and Energy Study Institute. Data Center Power Demands Are Contributing to Higher Energy Bills (2026). (Advocacy organization article, not peer-reviewed.) eesi.org/articles/view/data center-power-demands-are-contributing-to-higher-energy-bills

[13] PolitiFact. How much have data centers increased electricity prices? (2026). (Fact-check journalism, not peer-reviewed.) politifact.com/factchecks/2026/jun/12/elizabeth-warren/data centers-rising-electricity-costs

[14] Yue, D. & Zeng, Y. The Local Economic Effects of Data Center Entry. SSRN (2026). dx.doi.org/10.2139/ssrn.6497238

[15] CleanView v2 hyperscale data center dataset. Harvard/Toronto research collaboration (version dated 2026-10-07/2026-10-09). Internal snapshot, not publicly accessible. (Primary data source, not a publication.)

[16] U.S. Energy Information Administration. Form EIA-861, "Sales to Ultimate Customers," annual data files, 2021–2025. eia.gov/electricity/data/eia861 (Government data source; also the basis for this project's derived state- and utility-level retail-choice

shares.)

[17] Homeland Infrastructure Foundation-Level Data (HIFLD) / Oak Ridge National Laboratory. Electric Retail Service Territories. Frozen 2025-08-21 snapshot, mirrored via ArcGIS after the HIFLD portal's decommissioning. arcgis.com/home/item.html?id=597555ce8e4a4892a030784a7c657fdd (Government data source.)

[18] U.S. Bureau of Labor Statistics. Quarterly Census of Employment and Wages (QCEW), annual singlefiles, 2015–2024. bls.gov/cew (Government data source.)

[19] U.S. Bureau of Economic Analysis. Regional Economic Accounts, Table CAINC1 (“Personal Income Summary”), county-level bulk file, 2015–2024. apps.bea.gov/regional (Government data source.)

[20] Monitoring Analytics, LLC (PJM Interconnection's Independent Market Monitor). State of the Market Report for PJM, annual volumes 2022, 2024, and 2025 (zonal LMP data, 2021–2025). monitoringanalytics.com/reports/PJM_State_of_the_Market (Market monitor report, not peer-reviewed.)

[21] U.S. Census Bureau county boundaries (public-domain GeoJSON), via Plotly's public datasets repository. github.com/plotly/datasets (Public geographic reference data.)

[22] International Energy Agency. Energy and AI (2025). iea.org/reports/energy-and-ai (Intergovernmental agency report, not peer-reviewed.)

[23] Gour, N. Health implications of the rapid rise of data centers in Virginia: an exploratory assessment. Frontiers in Climate (2026). doi.org/10.3389/fclim.2026.1648912 (Peer-reviewed mini-review.)

---

1. The post-opening rise comes from two designs: a narrower stacked comparison (+3.5% to +7.7% in the first two years) and a broader Callaway and Sant’Anna estimator (rising toward about +24% by year seven). The broader design also found treated counties already diverging from comparison counties five years before any campus opened, a pre-existing trend that violates the parallel-trends assumption a causal reading requires, which is why this result is flagged as unconfirmed rather than treated as evidence of a real effect.↩