Is AI Trained on Public Money? Evidence from US Data Centers*

Adam Feher† UNIL

Emilia Garcia-Appendini‡ Norges Bank, UniSG, SFI

Roxana Mihet§ UNIL, SFI, CEPR

August 27, 2025

Abstract

Data centers now account for a rapidly growing share of U.S. electricity demand. We link UPS-measured load for 3,830 facilities to county retail tariffs, firm outcomes, and emissions for 2010-2023 and estimate continuous-treatment difference-in-differences. We find no discernible local spillovers: household and commercial prices, nearby firms' ICT spending, revenue, and employment, and county CO_2 emissions remain flat—consistent with cost-of-service regulation and planning absorbing recent growth. Calibrated to these pre-2023 nulls, a small RTO/utility incidence model shows that even a tenfold AI/data-center shock leaves household bills broadly unchanged when operators self-procure via PPAs and pass-through is muted; bills become material only when load is grid-sourced and regulators both pass through wholesale changes and allocate network upgrades to households. Incidence is therefore a matter of regulatory design rather than technological inevitability.

Keywords: AI, energy prices, spillovers, data centers, energy, electricity.

JEL-Codes: Q55, Q58, D24 , O33 , O44, L94.

*First version: December 9, 2024. We would like to thank Martin Brown, Boris Nikolov, Michael Rockinger, and Pari Sastry for thoughtful discussions on this project, as well as participants at CEBRA, MIT-MRS AI and Climate Risk, and EEA conferences for helpful feedback. Roxana Mihet acknowledges generous financial support from The Sandoz Foundation - Monique de Meuron program. The views expressed are those of the authors and do not necessarily reflect those of Norges Bank. All errors are our own.

†University of Lausanne. Contact: adam.feher@unil.ch.

‡Norges Bank, University of St. Gallen, and SFI. Contact: emilia.garcia@norges-bank.no.

§University of Lausanne, SFI, and CEPR. Contact: roxana.mihet@unil.ch.

1 Introduction

The digital economy increasingly runs on electricity, and at its center are data centers. Cloud computing, artificial intelligence training and inference, video streaming, and crypto-mining all rely on large facilities designed for uninterrupted operation. A single hyperscale campus can draw as much electricity as a mid-sized town, and collectively U.S. data centers already consume a sizable share of the national load. In 2023, these facilities accounted for roughly 4.4 percent of total electricity demand, with projections placing their share between seven and twelve percent by 2030.[1] Growth is driven not only by the expansion of cloud services but also by the rapid acceleration of AI. The International Energy Agency estimates that electricity demand linked to AI could increase tenfold between 2024 and 2026, virtually all of it channeled through data centers.[2]

The policy stakes are visible in places like Northern Virginia, where lawmakers debated whether unconstrained data-center growth would raise household bills absent reforms, and where private contracts with nuclear and renewable resources have become central to corporate procurement strategies.[3] These debates turn on spillovers: who ultimately pays: households, small businesses, or the operators themselves, and how that incidence depends on wholesale markets and cost-of-service retail regulation.

To adress this debate, we assemble a county-year panel that links the *measured* electricity load of 3,830 U.S. data-center facilities to local retail tariffs, firm outcomes, and emissions between 2010 and 2023. Using state-of-the-art *continuous* difference-in-differences estimators (with IV) and complementary event-study designs, we quantify the local spillovers of data-center demand on residential and commercial prices, nearby establishments' investment and employment, and county CO_2e emissions. We then build a small, institutionally grounded

[1] U.S. Department of Energy (2024).

[2] IEA (2024).

[3] See the Virginia JLARC study (2024) and coverage of nuclear-backed PPAs and plant preservation in industry reporting.

regional model that is calibrated to our pre-2023 estimates and U.S. tariff institutions, and we use it to stress-test forward scenarios consistent with projected AI-driven load growth. The model nests our historical null as a limit case and identifies conditions under which future load could change household bills or emissions.

On the data side, we convert facility-level UPS capacity and utilization from S&P Global's 451 Research *Data Center KnowledgeBase* into annual MWh by county, merge county tariffs from EIA Form 861, and match plant-level emissions from eGRID. For local real-activity outcomes we use the Aberdeen Technology Data Cloud (formerly Harte Hanks / SWZD), aggregating to county and commuting-zone levels.[4] On the empirical side, (i) we estimate continuous DiD models in reduced form and with an IV based on (a) state data–enter incentives, (b) county broadband availability and quality, and (c) distance to major hubs (e.g., Ashburn, VA); (ii) we conduct binary staggered-timing DiD and event studies as cross-checks; and (iii) we explore heterogeneity by treatment intensity.

On the modeling side, we specify a regional partial-equilibrium environment with an upward-sloping wholesale supply curve, a cost-of-service retail utility, and three primitives that determine incidence: wholesale pass-through, allocation of non-energy costs, and the share of DC load procured via PPAs. Calibration targets our pre-2023 empirical results and observed institutional features of the data centers and energy markets.

Our estimates show no detectable effect of data-center load on local residential or commercial tariffs, firm outcomes, or county CO_2e emissions through 2023. U.S. retail institutions and grid planning appear to have absorbed recent hyperscale growth without rate pressure on nearby households or businesses.

Where data centers expanded before 2023, household and business tariffs did not move in a systematic way; dynamic event studies are flat and continuous-treatment estimates with and without instruments corroborate the absence of price pressure. Nearby firms neither scaled up

[4] Sections **??**-4 detail construction, aggregation, and robustness.

nor contracted on average, IT budgets, employment, and revenues remain remarkably stable, consistent with the idea that hyperscale entry neither taxes local margins nor catalyzes broad commercial booms in the medium run. Local CO_2e emissions also do not rise in the counties that host facilities, in line with procurement strategies that source power system-wide and with the way U.S. retail institutions buffer short-run shocks. The simplest reading is institutional: cost-of-service regulation, rate design, and planning practices have so far absorbed data-center growth without shifting noticeable costs onto households or small businesses.

However, industry roadmaps and policy discussion, including prominent coverage of nuclear-backed PPAs and rapid AI load growth, suggest the next decade may not look like the recent past. Public debate now focuses on whether accelerating AI demand will raise local bills. Motivated by this gap between a historical null and a potentially different future, we develop a transparent regional model that is disciplined by our pre-2023 evidence and electricity-market institutions. The model reproduces our null under observed pass-through and cost allocation, and then asks: under what combinations of procurement strategy, pass-through, and allocation would tenfold DC demand begin to matter for household bills and system emissions?

We use the model to ask when rapid AI-driven growth could overturn the historical null. Three lessons emerge. First, procurement matters: when most DC energy is procured via PPAs, retail bills for households and small firms remain essentially insulated even under large load increases. Second, institutions dominate magnitudes: the same physical shock generates very different outcomes depending on the wholesale pass-through parameter and how new non-energy costs (capacity, T&D, interconnection) are allocated. High pass-through and grid-sourced growth shift the outcome from null to noticeable monthly bill additions. Third, the long run can exceed the short run even with similar wholesale prices because network and interconnection investments enter the rate base; the wedge is small if DCs finance a large share of upgrades and larger otherwise. These mechanisms reconcile a historical absence of spillovers with forward-looking concern: it is not only the megawatts that matter, but how the megawatts

meet the grid.

Our paper contributes to four different literatures.

First, we contribute to work on the economic geography and infrastructure footprint of data centers, documenting where load growth occurs and how it interacts with network planning and siting.[5]

Second, we speak to electricity-market incidence and regulation: we show why regulated tariffs can mute short-run pass-through while shifting non-energy costs to the rate base, complementing evidence on pass-through in competitive settings and classic insights on rate-of-return incentives.[6]

Third, we add to the emerging economics of AI and digital infrastructure by quantifying how compute-driven demand interacts with legacy utility institutionsâechoing industry reports on PPA strategies (including nuclear-backed contracts) and grid adequacy debates.[7]

Finally, methodologically, we deploy continuous-treatment DiD (and IV) with stayers and switchers and use modern staggered-timing estimators, bringing these tools to a setting where "how much" treatment matters at least as much as "whether" treatment occurs.

The rest of the paper proceeds as follows. Section 2 describes the data-center market and the U.S. electricity institutional setup. Section 3 presents our data. Section 4 details the empirical strategy. Sections 5 and 6 study price incidence on households and commercial customers. Section 7 turns to local firm outcomes, and Section 8 examines emissions. Section 9 develops the model and runs forward-looking scenarios. Section 10 discusses policy implications and concludes.

[5] See the overview of hyperscale clusters and siting economics in our introduction and references therein.

[6] Institutional background and references are summarized in Section 2, including market organization by RTO/ISO and the mechanics of cost allocation.

[7] See the discussion of PPAs, nuclear preservation efforts, and Loudoun County policy debates in Section 2.

2 Institutional Context

Data Centers. The role of data centers in the energy system has evolved over the past two decades. Early "enterprise" facilities, typically attached to corporate campuses, drew modest power relative to industrial users. Today, the largest hyperscalers, such as Amazon, Microsoft, Google, and Meta, operate clusters of facilities that each require continuous loads in the tens to hundreds of megawatts. Their siting decisions shape regional transmission planning, influence wholesale market outcomes, and increasingly enter policy debates about grid adequacy and decarbonization.

Data centers typically secure power through two channels. The first, still the dominant method, is to connect directly to the commercial grid under the prevailing tariff. This ensures stable and reliable electricity but also embeds the facility within the cost allocation rules that govern local utilities.[8] The second approach is to sign a long term private power purchase agreement (PPA) with independent generators. These contracts are often branded as "green" or "additional" capacity, linking corporate sustainability commitments with new renewable or nuclear supply.[9] In practice, PPAs shift the economics of siting decisions: they allow hyperscalers to bypass local utilities while still depending on the transmission system those utilities finance. Because hyperscale data centers can consume tens of megawatts continuously, their operators have also lobbied to preserve otherwise uneconomic nuclear plants as dedicated sources of baseload supply.[10]

The implications for host communities are contested. Proponents emphasize jobs in construction, tax revenues, and a cluster effect on local IT services. Critics highlight low permanent employment, high land use, and heavy energy and water footprints. In Virginia, where a substantial ratio of the world's internet traffic flows through Loudoun County, policymakers have

[8] Equinix Amazon

[9] See Google's 20-year PPA with Brookfield Hydro Brookfield; Microsoft's deal with Constellation Nuclear. Constellation Nuclear

[10] For example, Microsoft's lobbying to keep Three Mile Island operating. MIT Tech Review

debated whether unconstrained growth in data center demand risks higher retail electricity prices for households and small businesses. A legislative study estimated that, without reform, the average household bill could rise by $14-$37 per month by 2040 (Joint Legislative Audit and Review Commission, 2024).

Electricity Market Institutions Understanding the incidence of such growth requires an overview of U.S. electricity market institutions. Retail delivery remains dominated by state regulated monopoly utilities. Investor-owned utilities are granted exclusive service territories and earn a regulated rate of return on capital, approved in periodic rate cases. Wholesale trade is coordinated by federally supervised regional transmission organizations (RTOs) such as PJM, MISO, and ERCOT. These RTOs dispatch generation through auctions, allocate transmission and capacity costs to members, and set the marginal price of electricity in their markets. As a result, retail prices observed by households and firms reflect a combination of state level cost recovery decisions and federally organized wholesale outcomes (Martin and Peskoe, 2025).

Retail tariffs follow the cost-of-service model: a fixed charge, volumetric per-kWh rates, and demand charges tied to peak load. Utilities frequently add discounts or long-term contracts to attract energy-intensive facilities like data centers, with unrecovered costs spread across other ratepayers when regulators approve. Transmission upgrades and capacity-market obligations assigned by RTOs likewise flow into retail bills. Importantly, the allocation formulas that govern these charges are path dependent and highly technical. Yet their distributional consequences are large: seemingly minor changes in how transmission costs are shared can redistribute millions of dollars across residential, commercial, and industrial classes (Martin and Peskoe, 2025).

The economics of passthrough further clarifies why data centers matter. In competitive wholesale markets, input-cost shocks typically move prices almost one-for-one. For Spain, Fabra and Reguant estimate that a 1 euro increase in CO_2 permit costs raises wholesale electricity prices by roughly 0.80 euro on average, and nearly 1 euro during high-load hours. In the U.S.,

however, regulated retail tariffs mute short run responses but spread long run costs across the rate base. If a large new industrial user arrives, the utility's revenue requirement rises, unless special contracts capture the incremental cost, other customers eventually pay. Under rate-of-return regulation, this can even incentivize utilities to over-invest in capital (Averch and Johnson, 1962).

3 Data

3.1 Data centers

Our data on data centers come from S&P Global's 451 Research *Data Center KnowledgeBase*.[11] This comprehensive database covers 3,830 U.S. data center facilities, including hyperscale, leased, crypto, colocation, and wholesale centers. It provides both point-in-time snapshots and historical time series of facility- and market-level capacity and utilization, thereby enabling analysis of past, present, and future data center growth.[12]

Crucially, the database reports each center's total utilized UPS power for every year from 2017 through 2023, which we convert to annual energy consumption. We aggregate this consumption to the county level and compare it with overall county energy use. Over this seven year span, data center energy consumption has roughly doubled, closely mirroring the rise in total commercial energy drawn from the grid.[13]

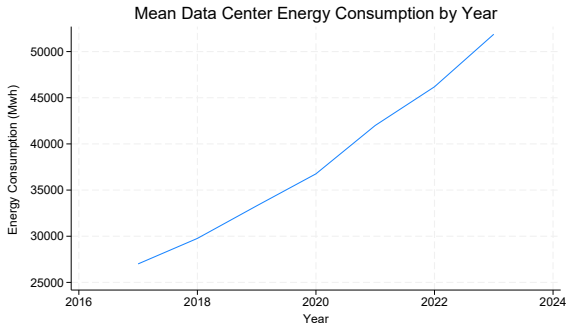
Figure 2 highlights the data center energy load for each county. Darker colors indicate higher loads. Overall, the map shows that there are larger data centers close to the economically most active areas, such as East-coast and West-coast hubs, large tech cities in Texas, etc. This also coincides with the observations made in Greenstein and Fang (2020)

[11] S&P Global

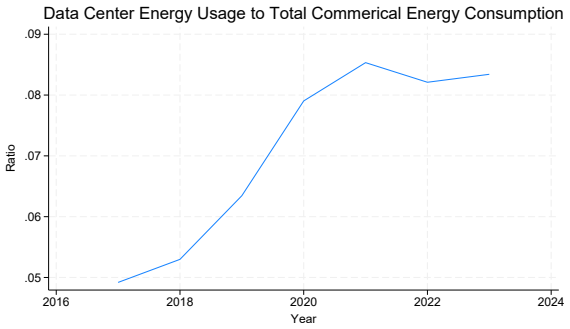
[12] The database includes centers that closed by 2024 as well as planned facilities that are not yet fully operational.

[13] Some of this power may be procured via private power purchase agreements (PPAs).

Figure 1: Data center energy consumption (absolute (left) and relative (right))



Absolute data center energy consumption



(a) Relative to total commercial consumption

3.2 Energy data

Electricity prices and quantities are drawn from the U.S. Energy Information Administration or EIA.[14] In particular, FERC form 861 reports cents per kWh charges and MWh sales for every investor-owned, municipal, and cooperative utility. The prices are shown on the following heatmap. Electricity prices are generally higher in areas with more economic activities, such as along the two coasts of the country.

Moreover, we access power plant level information, such as the emission of each plant for the years 2016-2023, via the EGRID website of EIA. [15]

3.3 Aberdeen

In the second part of the paper, we analyze how firms respond to the elevated energy demand from data centers. Our primary source is the Aberdeen Technology Data Cloud, formerly known as Harte Hanks (HH) and Spiceworks Ziff Davis (SWZD).[16] The database provides annual financial and IT metrics for millions of U.S. establishments between 2010 and 2022.

[14] EIA

[15] EIA-EGRID

[16] The survey methodology changed over time, likely due to the change in ownership. We have also noticed changes in our dataset, especially inconsistencies of certain observations between years. For that reason, we always use the Aberdeen data aggregated at the county/commuting zone level, moreover our method

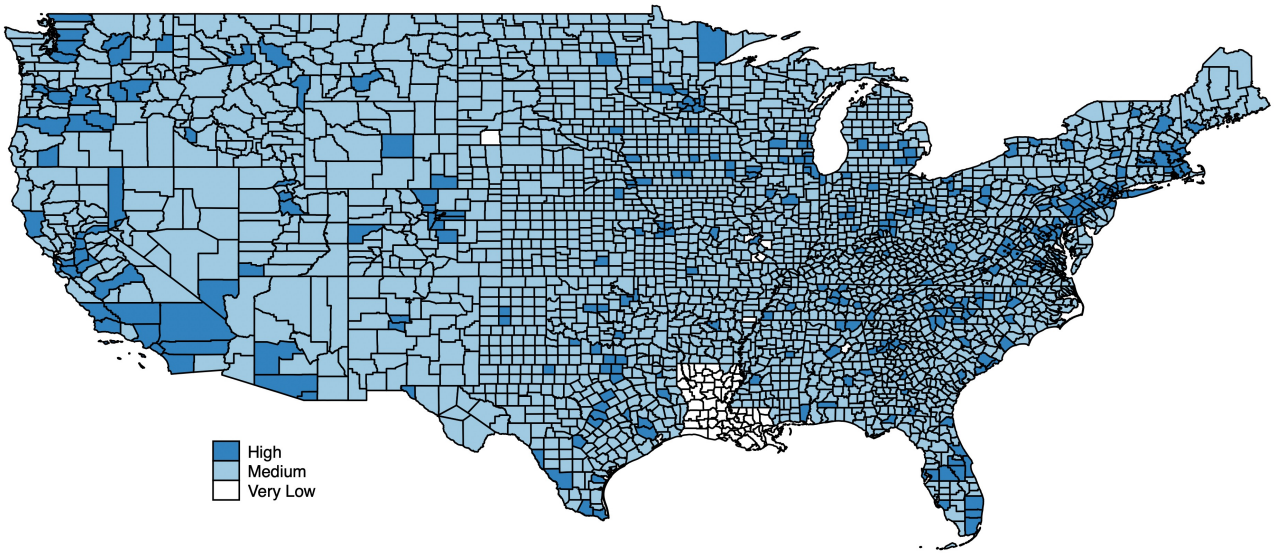


Figure 2: Data center energy consumption at the county level

Darker shading indicates higher electricity use (MWh). Values are aggregated to the county–year level for 2017. Counties without data centers are shown in white.

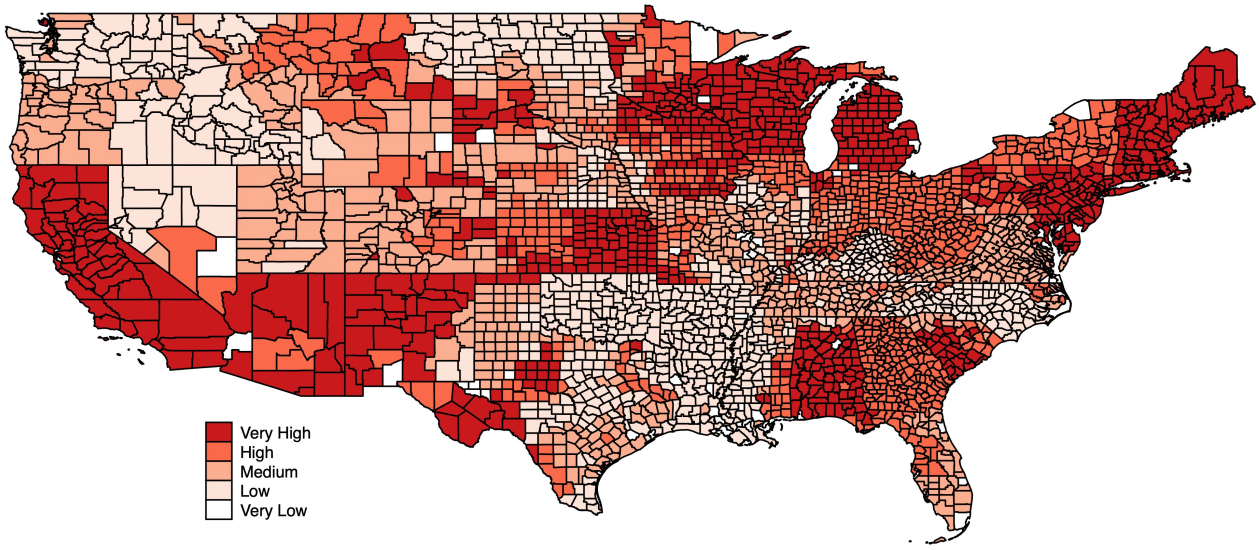


Figure 3: Energy prices heatmap

Darker shading indicates higher electricity prices. Values are aggregated to the county-year level for 2017.

From this database, we construct our key outcomes: ICT spending, revenue (in USD), and both IT-specific and total employment at the establishment level.

3.4 Control variables

We use a diverse set of control variables to capture variation in data center energy consumption, location, and trends on the energy markets.

3.4.1 Incentives to attract Data Centers

Many states have implemented incentives to attract new data centers. We summarize these in a binary dataset, indicating whether in a given year a state has or does not have any enacted policy to support data centers, based on the survey carried out in Husch Blackwell (2025).

3.4.2 Broadband quality

Broadband, or internet reliability and speed, is a crucial infrastructural need of data centers. The Federal Communications Commission (FCC) maintains a granular database of broadband deployment through out the United States. We use several measures to aggregate this data to county level. [17]. The data is publicly available on the FCC's website. [18]

3.4.3 Outside air temperature

Temperature is a relevant variable both for the data centers but also for the broader energy markets. If the temperature is cold enough, the machines in the center might be cooled simply by outside air. The air temperature is also relevant for the electricity market and prices, since in cold households might heat, while at the warmest temperature they may use air conditioning.

[17] We need to aggregate the information given at the census block level to the larger county level. We use several measures, such as the download speed and the upload speed, and the transmission quality, ie.: Fiber, DSL etc.

[18] FCC Broadband Data

We obtain temperature data based on the average temperature in the 20th century, form the National Centers for Environmental Information of the National Oceanic and Atmospheric Administration. [19]

3.5 Macro variables

We use US census data provided by the Bureau of Economic Analysis (BEA) for county GDP, territorial size, and population. We utilize the National Bureau of Economic Research (NBER) data to calculate geographical distances between counties and cities. [20] [21] We obtain the median home prices from the National Association of Realtors (NAR), and an index for the risk of natural disasters by the Climate Mapping for Resilience & Adaptation (CMRA) database. [22] [23]

Table 1: Summary of databases used

Database	**Description**
Aberdeen	ICT and revenue data at establishment level
Energy Information Administration (EIA)	Energy production per state
Database of State Incentives for Renewables & Efficiency (DSIRE)	Policies and incentives for renewable energy

3.6 Constructing county-level quantities and prices

3.6.1 Energy

The geographical unit of our analysis is the county level at the United States. Therefore, we need to transform our data to the county level. ERC Form 861 lists each investor-owned, municipal, and cooperative utility's annual kilowatt-hour sales and its average charge in cents per kWh. Because a utility often supplies several counties and most counties are served by

[19] NCEI temperature data
[20] BEA
[21] NBER county distances
[22] NAR home prices
[23] CMRA risk data

more than one utility, a county-wide tariff cannot be observed directly. We therefore construct three complementary price measures. First, we compute an unweighted ("simple") average of the tariffs posted by every utility that operates in a given county; this tends to put more weight on small providers. Second, we calculate a sales-weighted average that assigns to each utility a weight proportional to its total national sales; if larger utilities indeed cover larger shares of a county's load, this delivers a closer approximation to the true county price, although the weight is exact only under that proportional-share assumption. Third, we restrict the sample to counties in which exactly one utility operates, thereby eliminating the cross-utility aggregation problem altogether. Hence, the remaining measurement error stems only from the possibility that the same utility charges different prices in neighboring counties. Applying these three strategies yields county-year series for both commercial, and household electricity prices that we use throughout the analysis.

3.6.2 Data center demand

For each facility we observe the UPS (critical) capacity in MW and an annual utilization factor. The product of the two, annualized, gives us the yearly data center energy consumption in MWh. Summing across all facilities located in county yields a yearly measure of data center energy consumption. About 80% of US counties have not had a data center until 2023, which group serves as the control group for our estimation.

4 Methodology

4.1 DiD estimation techniques

Our primary method for estimating the causal impact of data centers' energy consumption on energy markets and the broader economy is the continuous difference-in-differences (DiD) approach developed by de Chaisemartin and D'Haultfoeuille (De Chaisemartin and D'haultfoeuille,

2020; de Chaisemartin and D'Haultfœuille, 2020; De Chaisemartin et al., 2025). Unlike binary treatment DiD designs, which only capture whether a data center is present or not, the continuous DiD framework allows us to directly incorporate variation in energy load. This is crucial because a modern hyperscale facility operating at > 100 MWh is likely to affect markets differently than a smaller data center drawing < 10 MWh. Moreover, the framework accommodates instrumental variables, enabling us to address potential endogeneity in both location and load size choices.

The continuous DiD approach assumes that between consecutive periods some units - the *switchers* - change their treatment, while others - the *stayers* - do not. Under a parallel trends assumption, weighted averages of the switchers' potential outcome slopes are nonparametrically identified by difference-in-differences estimands that compare outcome trajectories of switchers and stayers with the same baseline treatment level. Conditioning on the baseline treatment ensures that the estimands remain valid even if the treatment effect evolves over time (De Chaisemartin et al., 2025).

More formally, the crucial parallel trends assumptions with continuous treatment reads as follows. Let D_1 (respectively D_2) denote the unit's treatment in period 1 (respectively 2), and let $\mathcal{D}_1$ (respectively $\mathcal{D}_2$) be its support. Define the indicator $S = \mathbf{1}\{D_2 \neq D_1\}$, which equals 1 when the unit's treatment changes between periods 1 and 2, that is, when the unit is a switcher. For any $d \in \mathcal{D}_1 \cup \mathcal{D}_2$, let $Y_1(d)$ and $Y_2(d)$ denote the unit's potential outcomes in periods 1 and 2 under treatment d, and let Y_1 and Y_2 be the corresponding observed outcomes. Our potential outcome notation assumes that Y_1 does *not* depend on the unit's period-two treatment, thereby ruling out anticipation effects, a commonly made assumption in the difference-in-differences literature (De Chaisemartin et al., 2025).

$$\forall\, d_1 \in \mathcal{D}_1 :\ \mathbb{E}[\Delta Y(d_1) \,|\, D_1 = d_1,\, D_2] \ =\ \mathbb{E}[\Delta Y(d_1) \,|\, D_1 = d_1]\,. \tag{1}$$

In other words, the parallel trends assumption requires that $\Delta Y(d_1)$ be mean independent of D_2, conditional on $D_1 = d_1$.

Our parameters of interest are the Average Slope (AS) and Weighted Average Slope (WAS).

The AS is defined as

$$\delta_1 \;:=\; \mathbb{E}\left[\frac{Y_2\Big(D_2\Big) \;-\; Y_2\Big(D_1\Big)}{D_2 - D_1} \;\middle|\; S = 1\right]. \tag{2}$$

AS is a local effect that applies only to switchers. It measures the effect of changing their treatment from its period-one to its period-two value.

The WAS is defined as

$$\delta_2 \;:=\; \mathbb{E}\left[\frac{|D_2 - D_1|}{\mathbb{E}[|D_2 - D_1| \,|\, S = 1]} \times \frac{Y_2(D_2) - Y_2(D_1)}{D_2 - D_1} \;\middle|\; S = 1\right]. \tag{3}$$

δ_2 is a weighted average of the slopes of switchers' potential outcome functions from their period-one to their period-two treatments, where each slope receives a weight proportional to the switcher's absolute treatment change between the two periods.

The AS and WAS serve different purposes. Unlike the AS, the WAS cannot be used to identify or bound the effect of treatment changes other than those that actually occurred, even under shape restrictions on the potential outcome function. Instead, the WAS may be employed to conduct a cost-benefit analysis of the treatment changes that took place from period 1 to period 2 (De Chaisemartin et al., 2025).

It is also possible to extend the above method to more than two periods via the appropriate weighting. In the case of AS, the weights correspond to the share of switchers from $t-1$ to t. With WAS, the weights are proportional to the average absolute treatment change from $t-1$ to t, yielding a WAS estimator for more than two periods. Those estimators are consistent if (t–1)-to-t switchers and stayers with the same (t–1) treatment would have experienced the same average ($t-1$)-to-t outcome evolutions if switchers had not switched. We assess the plausibility

of this assumption via "pre-trends" (or "placebo") estimators (De Chaisemartin et al., 2025).

We also use binary DiD specifications as a robustness check. In this case, we record the year each data center opened and code treatment as a binary indicator that switches from 0 to 1 in the opening year and remains 1 thereafter. Because treatment timing is staggered, some centers open earlier than others, we apply methods specifically designed for this case, such as Borusyak et al. (2024); Callaway and Sant'Anna (2021). However, these estimates serve only as a robustness check; our primary methods are the continuous DiD specifications.

5 Data centers' impact on residential electricity prices

The central question in this section is whether the rise in data centers' energy demand has affected residential electricity prices. To address this, we employ two complementary difference-in-differences (DiD) approaches: a continuous design based on actual energy load, and a binary design that codes a county as treated once it hosts a data center.

5.1 Contiuous DiD

We begin by examining the overlap assumption for the continuous DiD. Specifically, no value of D_1 should be observed only among switchers. While it is not feasible to verify this for every D_1, the distribution's mass point is at 0. In practice, about 80% of counties never hosted a data center up to 2023. During our observation window (2017-2023), however, new data centers entered each year in a subset of counties, ensuring that all such observations can be included in the estimation.

Table 2 shows that both the average slope and the weighted average slope are not statistically different from zero. In other words, data center energy load has no measurable effect on residential electricity prices. The table also highlights that only about 5% of observations are switchers, while the rest are stayers. This pattern arises because most counties never host a

data center and thus remain stayers at 0. At the other end, many data centers operate at full capacity, meaning that some counties with data centers also fall into the stayer category.

Table 2: Effect of data center energy load on residential prices (Continuous DiD, reduced form)

Number of observations: 18,231	**Method:** doubly-robust		**Polynomial order:** 1			
	Estimate	SE	LB CI	UB CI	Switchers	Stayers
Average Slope (AS)	0.0005844	0.0016361	-0.0026	0.0038	1,107	17,124
Weighted Average Slope (WAS)	-4.84e-06	9.98e-06	-0.0000	0.0000	1,107	17,124

The table shows the reduced from estimates of AS and WAS, including time varying control variables (GDP, population). Standard errors are clustered at the county level. We find no significant effect of the increased data center energy consumption on residential prices.

A crucial assumption of the method is that switcher and stayers follow a common pre-trend that can be tested via placebo estimates. The placebos are analogous to the actual estimates, but they replace ΔY_t by ΔY_{t-1}, and they restrict the sample to counties whose data center energy consumption did not change between $t-2$ and $t-1$. Table 3 shows that our placebo estimates, both for AS and WAS, are statistically insignificant.

Table 3: Placebo tests for the effect of data center energy consumption on residential prices (Continuous DiD design, reduced form)

Number of observations: 14,197	**Method:** doubly-robust		**Polynomial order:** 1			
	Estimate	SE	LB CI	UB CI	Switchers	Stayers
Placebo_1 (AS)	0.0003584	0.0014197	-0.0024242	0.0031409	49	14,148
Placebo_1 (WAS)	-4.30e-07	8.98e-07	-2.19e-06	1.33e-06	49	14,148

The table shows that the placebo or pre-trend estimates are 0, hence we are confident about the parallel trend assumption.

We also test the robustness of our result by removing our control variables, changing the treatment's functional form to be logarithmic, and winsorizing the treatment. The latter two robustness checks are primarily to account for the highly skewed distributions of data center energy consumption; some counties have many large data centers, while most counties do not have any. Table 4 highlights that our null results are robust to these changes.

Table 4: Effect of data center energy consumption on residential prices (Continuous DiD design, reduced form)

	No Controls	**Log energy usage**	**Winsorized energy usage**
Average Slope (AS)	0.0028228 (0.196169)	197.3225 (233.5715)	0.0006248 (0.0016683)
Weighted Average Slope (WAS)	-0.005421 (0.0107347)	-0.6990835 (0.7430867)	-0.0000105 (0.0000114)
# Switchers	1,111	1,090	1,078
# Stayers	17,119	722	17,152

The table presents three robustness checks that confirms our previous null result, that increased data center energy usage does not drive up residential prices. The first column removes control variables, the second uses logarithm of data center usage, and the third column winsorizes at 0.2%, approximately corresponding to the top 1% of counties that do have a datacenter(s).

5.2 Continous IV-DiD

An additional advantage of the continuous treatment is that it allows us to address potential endogeneity in data centers' energy loads by instrumenting the treatment. We employ three instruments. The first captures state-level regulatory incentives for building and operating data centers. The second reflects broadband availability and quality at the county level. The third is an interaction term measuring a center's distance from Ashburn in Loudoun County, widely recognized as the largest US data center hub, conditional on the center not being adjacent to another major city hub. The relevance of the distance to Ashburn and other major cities is also emphasized by Greenstein and Fang (2020).

Since the estimation package developed by De Chaisemartin et al. (2025) only permits a single instrument, we use the first principal component of the three variables listed above. As before, we begin by checking the overlap assumption. Figure 4 presents the kernel densities based on data from 2017. It highlights that the distributions of switchers and stayers are similar, hence we are confident again that the overlap assumption is satisfied.[24]

The first row of Table 5 shows that the first stage is strong. The second row indicates that our IV estimate is very similar in magnitude to the reduced form result and is likewise

[24]In fact, at p=0.01 we cannot reject that the two distributions are identical. The distributions between other pairs of years are also similar.

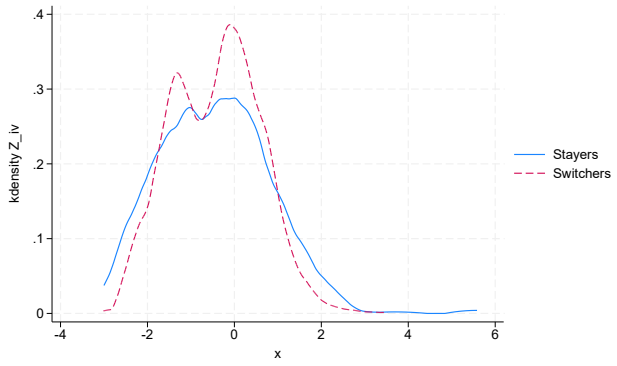


Figure 4: Kernel density of instrument with the base year 2017

insignificant. We also reestimate the model using bootstrapped standard errors instead of the theoretically derived ones from De Chaisemartin et al. (2025). The results remain nearly identical in this case as well.

Table 5: IV estimates (Continuous DiD design)

Number of observations: 18,217			**Method:** doubly-robust		**Polynomial order:** 1	
	Estimate	SE	LB CI	UB CI	Switchers	Stayers
First stage WAS	8,654.72	3,138.244	2,503.762	14,805.68	16,087	2,130
IV-WAS	4.67e-06	0.0002741	-0.0005325	0.0005418	16,087	2,130

Notes: Our first stage shows that the instrument is statistically significant. In the second row we report the IV estimate that is not significant. Hence, the energy load of data centers do not impact residential electricity prices.

In the following sections, we conduct additional robustness checks using alternative DiD and event study methods with a binary treatment.

5.2.1 Binary treatment

The binary treatment indicates whether a county hosts a data center or not. To ensure comparability across facilities, we restrict the analysis to the years 2010-2023.[25] This time window is therefore broader than in the continuous treatment analysis.

[25] Comparing the impact of a 1980s data center with that of a 2020s hyperscaler would likely yield unrealistic results.

Table 6: First-stage and IV estimates of the Weighted Average Slope (Bootstrap SE for IV-WAS)

Number of observations: 18,217		**Method:** doubly-robust		**Polynomial order:** 1	
	Estimate	SE	LB CI	UB CI	Switchers	Stayers
First stage WAS	8,654.72	3,138.244	2,503.762	14,805.68	16,087	2,130
IV-WAS	4.67e-06	0.0037468	-0.0002787	0.0004606	16,087	2,130

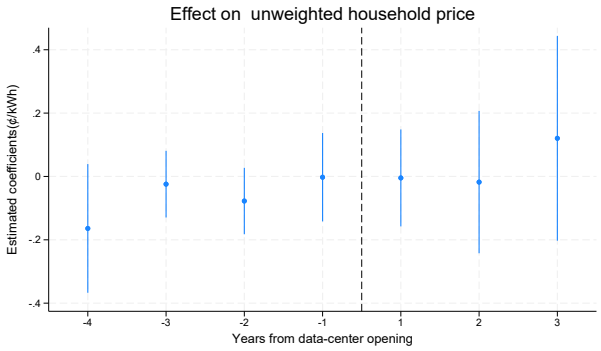
The table reports the IV results based on bootstrapping with n=100. The bootstrapped and theoretically calculated standard errors are similar. Our coefficient of interest statically insignificant as it was in our previous estimations.

Our empirical strategy for the binary treatment has two complementary parts: *(i)* an event study that traces the full path of prices before and after entry, and *(ii)* a difference-in-differences (DiD) estimate of the average treatment effect once a center is in operation.

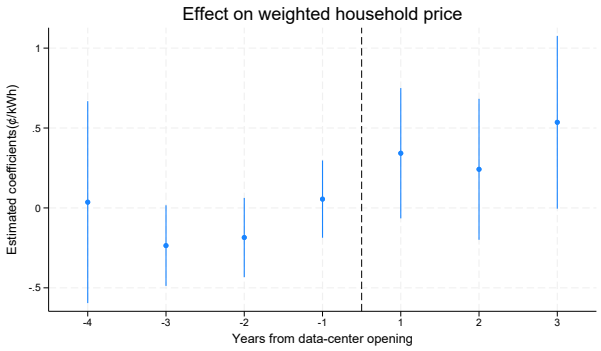
5.2.2 Event-study evidence

Figure 5 plots dynamic coefficients from the Callaway-Sant'Anna estimator using a four-year window on either side of entry. Panel A shows the unweighted residential price, Panel B the sales-weighted price, Panel C the residential-commercial spread, Panel D the year-over-year change, and Panel E the weighted price in single utility counties.[26] In every specification the four leads (years -4 to -1) are not significantly different from zero ($p > 0.10$), indicating that treated and control counties share parallel trends prior to the first data center coming online. After entry, the point estimates remain clustered around zero and none is statistically different from the reference year at conventional significant levels. Visually, the price path is flat. The event study therefore delivers a clear message: conditional on county and year fixed effects, the arrival of a data center does not move household electricity prices.

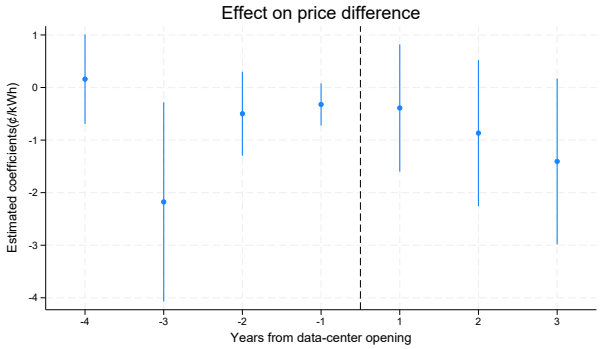
[26] Section 3.6.1 describes the creation of the different outcome proxies.



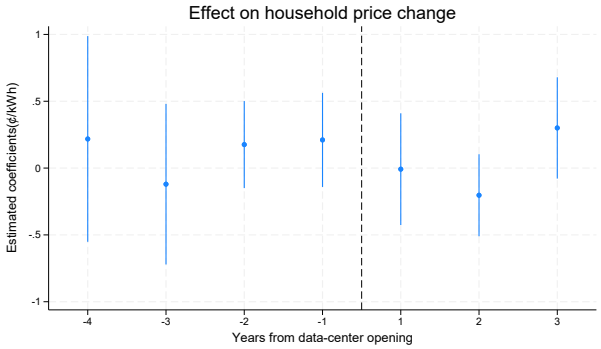
(a) Unweighted household price



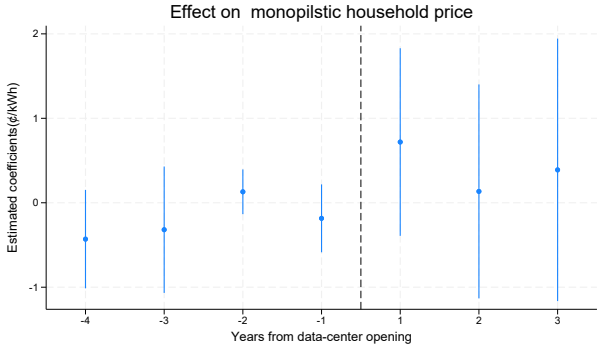
(b) Utility sales weighted household price



(c) Price difference between residential and commercial price



(d) Panel D



(e) Monopoly utility provider

Figure 5: Pre-trend of event study

Each point is the difference-in-differences ATT coefficient for the indicated event time bin; vertical bars are 95 % confidence intervals clustered at the county level. The dashed vertical line marks the year of the first data center opening in a county. Panel A uses simple averages, Panel B sales weighted averages, Panel C annual price changes, and Panel D the subset of counties served by a single utility. The graphs show that the parallel trend assumptions are satisfied.

5.2.3 Binary staggered DiD

To exploit the longer 2010-2023 window we collapse treatment to a binary indicator that turns on in the first year a county hosts a data center and remains one thereafter. Applying the Callaway-Sant'Anna estimator with the 2,500 never treated counties as controls, Table 11 reports the average treatment effect on the treated. Across four of the five columns the ATT is small insignificant. Only column 2-where prices are weighted by utilities' national sales-yields a positive and significant effect of 0.8 cent/kWh roughly $7 per month for the median household. Because that weighting scheme exaggerates the influence of very large providers, we view this as an upper bound; the unweighted and single-provider specifications show no effect.

Table 7: Data Centers' Impact on residential price (Binary, Staggered DiD)

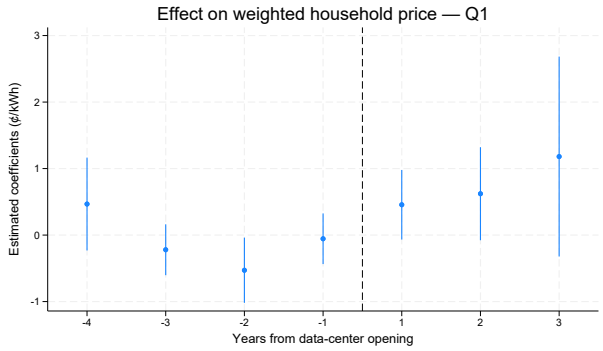
	Price_{uw}	Price_{w}	Price_diff	Δ price	Monopoly Price
	(1)	(2)	(3)	(4)	(5)
ATT	0.248	0.779^{**}	-0.028	0.262	0.149
	(0.176)	(0.383)	(0.134)	(0.191)	(0.648)
County FE	✓	✓	✓	✓	✓
Time FE	✓	✓	✓	✓	✓
Observations	21,760	21,760	21,760	21,746	5,584

Robust standard errors clustered by county in parentheses. $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$.

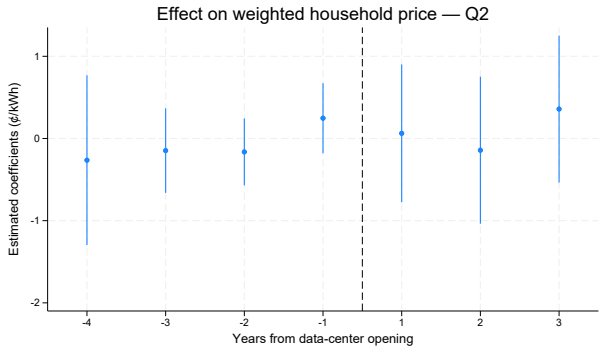
5.2.4 Multilevel staggered DiD

We bin our treatment variable into 4 intensity groups based on the year with the largest energy consumption by data centers in the county. That is, counties with the smallest, but positive, energy load are in the first quartile (q1), with the largest ones belonging to (q4). We continue to use the never treated counties as the control group. Moreover, to maintain our sample size, we make the assumption that the counties are not switching from one quartile to the other, once they are treated. That means, we extend our bins to years we do not actually observe the energy consumption of data centers. We focus on the weighted household price. Figure 6 shows that there is no heterogeneous impact based on the size of the data centers in a county.

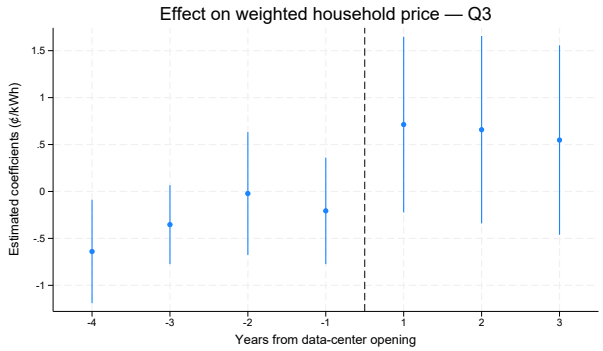
That is, even the counties with the largest overall energy load do not experience a significant increase in the residential energy price.



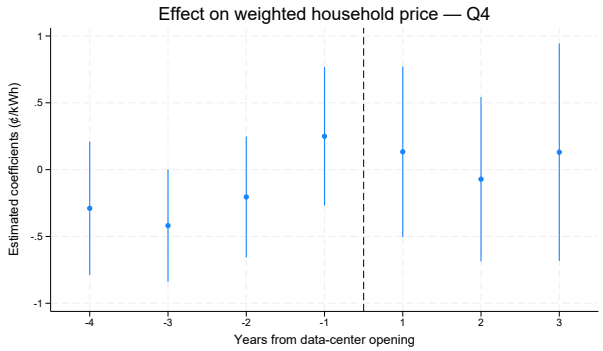
(a) Counties in the first quartile



(b) Counties in the second quartile



(c) Counties in the third quartile



(d) Counties in the fourth quartile

Figure 6: Multi level treatment

Each graph presents the event study with a multi-level treatment, where treatment intensity is divided into quartiles based on the maximum dosage between 2017 and 2023. The results show that data centers have no significant impact on residential electricity prices.

6 Commercial electricity prices

Next, we turn to the tariffs charged to commercial entities This is also the category data centers themselves belong to. [27] All specifications replicate the residential price analysis but replace the outcome with the county-year commercial price data.

[27] EIA

6.1 Continuous DiD

Similar to the case of residential prices, we find that the increased electricity load of data centers does not significantly raise commercial prices, as reported in Table 8.

Table 8: Effect of data center energy load on commercial prices (Continuous DiD, reduced form)

Number of observations: 18,278	**Method:** doubly-robust		**Polynomial order:** 1			
	Estimate	SE	LB CI	UB CI	Switchers	Stayers
Average Slope (AS)	0.0044956	0.0048196	-0.00495	0.01394	1,111	17,167
Weighted Average Slope (WAS)	-6.56e-06	0.0000106	-0.00003	0.00001	1,111	17,167

Notes: The reduced form estimates show no significant increase in the commercial price of energy.

Furthermore, the pre-trend assumption is satisfied, and the two-stage IV estimate also confirms the null result. These findings are reported in Tables 9 and 10.

Table 9: Placebo tests for reduced form

Number of observations: 18,278	**Method:** doubly-robust		**Polynomial order:** 1			
	Estimate	SE	LB CI	UB CI	Switchers	Stayers
	Estimate	SE	LB CI	UB CI	Switchers	Stayers
Placebo_1 (AS)	-0.0014483	0.0016049	-0.0045939	0.0016973	49	14,194
Placebo_1 (WAS)	-6.77e-07	1.78e-06	-4.17e-06	2.82e-06	49	14,194

Notes: The placebo tests are in alignment with our pre-trend assumption.

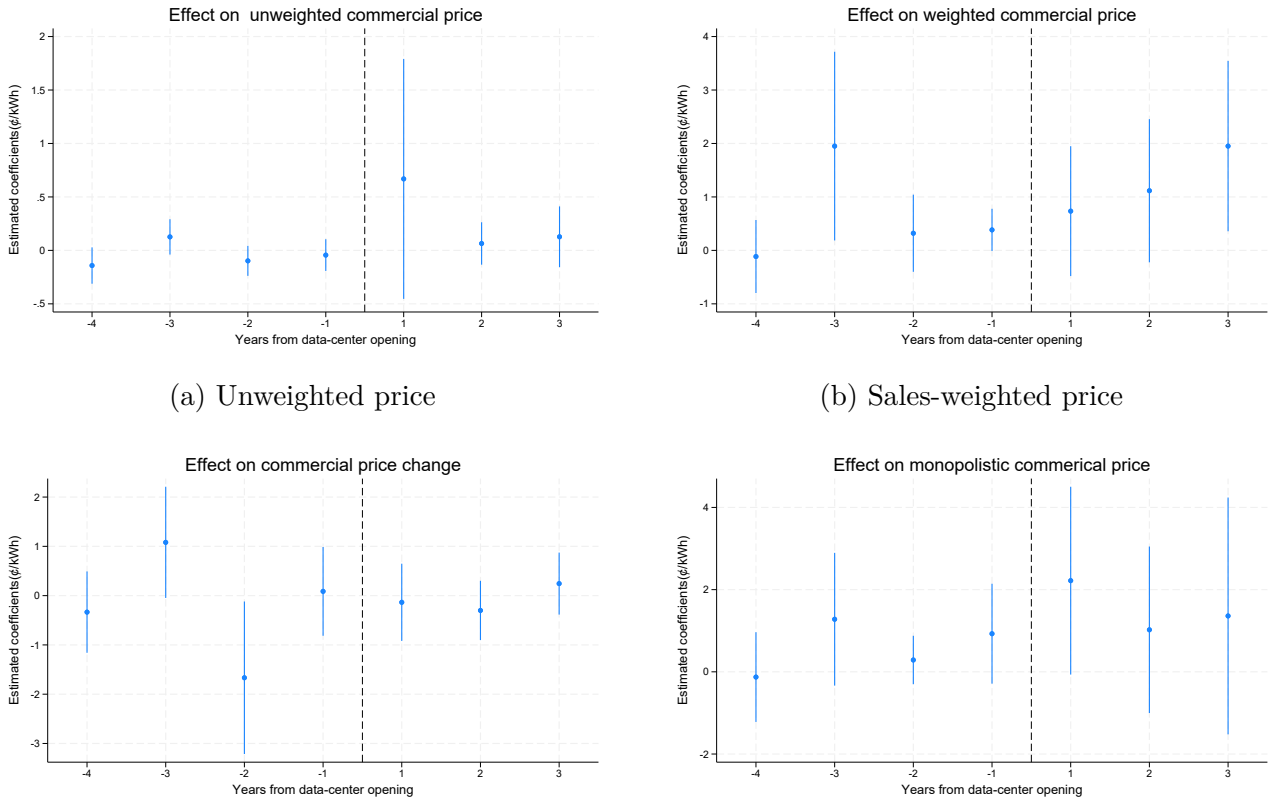
Table 10: First-stage and IV estimates of the Weighted Average Slope

Number of observations: 18,265	**Method:** doubly-robust		**Polynomial order:** 1			
	Estimate	SE	LB CI	UB CI	Switchers	Stayers
Weighted Average Slope (WAS)	8,686.159	3,139.117	2,533.491	14,838.83	16,121	2,144
IV-Weighted Average Slope (IV-WAS)	-0.0001375	0.0006767	-0.0014638	0.0011888	16,121	2,144

Notes: The first stage shows a statistically significant instrument. The second row present the the IV estimate that are in alignment with the reduced form results.

6.2 Binary treatment

Panel A-E of Figure 7 plots the event study coefficients obtained with the Callaway-Sant'Anna estimator. The four pre-treatment years coefficient are not significantly different from zero, supporting the parallel-trends assumption. Every post-entry coefficient is small and not significant.



(a) Unweighted price

(b) Sales-weighted price

(c) Year-over-year change

(d) Monopoly utilities only

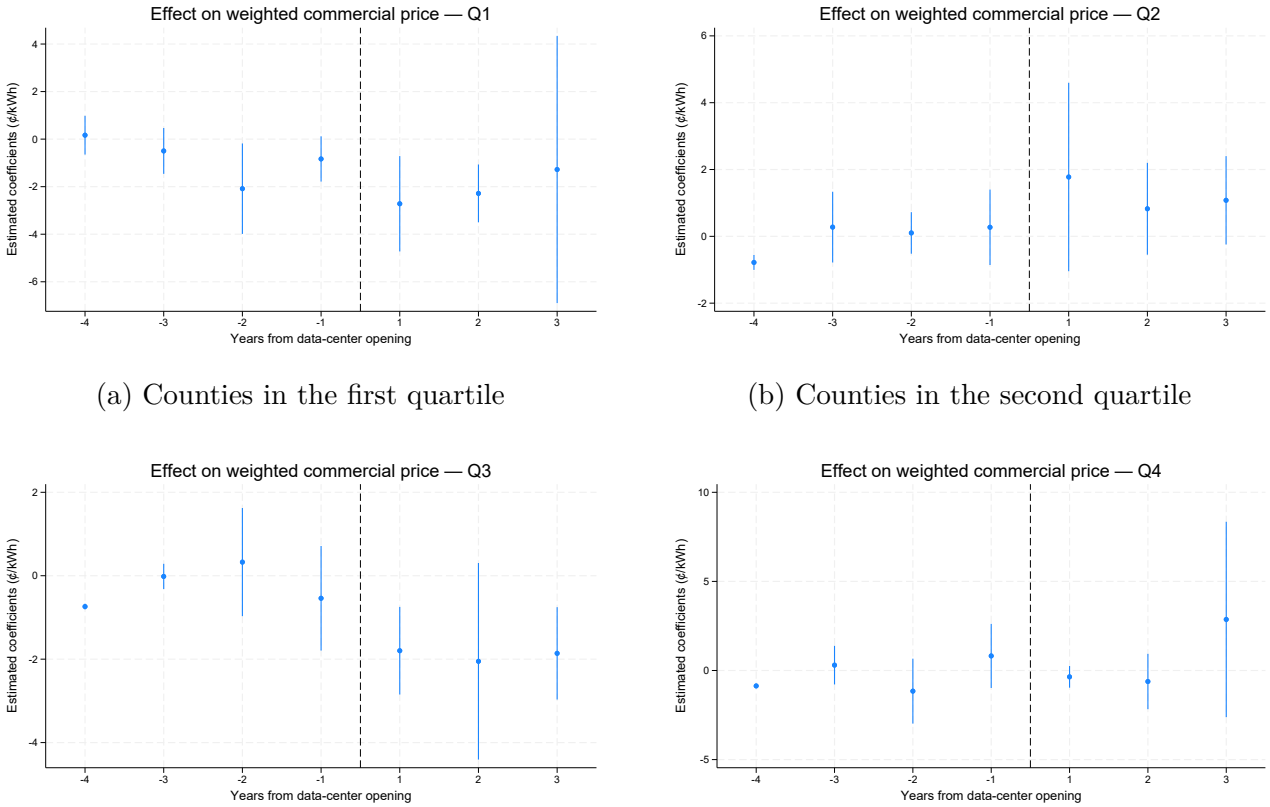
Figure 7: Event-study estimates for commercial electricity prices
Each point is the difference-in-differences ATT coefficient for the indicated event time bin; vertical bars are 95 % confidence intervals clustered at the county level. The dashed vertical line marks the year of the first data center opening in a county. Panel A uses simple averages, Panel B sales weighted averages, Panel C annual price changes, and Panel D the subset of counties served by a single utility. The graphs show that the parallel trend assumptions are satisfied.

Extending the horizon back to 2010 with the binary treatment yields the ATT reported in Table 11. As with households, the ATT is statistically insignificant in all specifications except

column 2, which shows a 2.58 cents/kWh increase in the tariff. This increase would translate into an additional cost of roughly 260 USD per year for a representative establishment consuming 100 MWh of electricity.

6.3 Multi level treatment

Considering heterogeneous treatment effects, we find that commercial electricity prices decline in counties with the smallest data center energy load, as shown in Figure 8. The ATT for the first quartile is -1.5 cents/kWh, which translates into roughly a 150 USD lower utility bills for an establishment in a county with relatively low data center energy demand compared to one in a county without any data centers.



(a) Counties in the first quartile

(b) Counties in the second quartile

(c) Counties in the third quartile

(d) Counties in the fourth quartile

Figure 8: Multi level treatment commercial

Each graph represents the event study based multi level treatment where treatment is binned into quartiles based on the max dosage between 2017 and 2023.The graphs show that data centers binned into the first quartile significantly decrease the commercial electricity prices. Other quartiles do not significantly impact the commercial price.

Table 11: Data Centers' Impact on Commercial Price (Binary Staggered DiD)

	Price_{uw}	Price_{w}	Δ price	Price (mono.)
	(1)	(2)	(3)	(4)
ATT	0.2867	2.5803**	−0.0416	0.7446
	(0.1747)	(1.0740)	(0.2423)	(1.3778)
County FE	✓	✓	✓	✓
Time FE	✓	✓	✓	✓
Observations	21,760	21,760	21,746	5,584
R^2	0.9712	0.9347	0.0670	0.9700

Robust standard errors clustered by county in parentheses.
$^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$.

7 Impact on the local economy

We now turn to the effects of data center entry on nearby IT-related establishments. Our analysis focuses on three core measures of firm level activity: ICT spending, total revenue, and employment. Because the scale of a data center's energy load likely matters for neighboring firms, we employ the continuous difference-in-differences specification only. A large facility may stimulate local IT investment by raising demand for cloud, storage, and related services. At the same time, it could bid away skilled labor from existing firms. The net effect is therefore theoretically ambiguous, depending on the relative strength of these opposing channels.

Table 13 presents the estimation results. Across all three outcomes, the coefficients are statistically insignificant. These null results align with the earlier evidence on no increase of the retail energy prices and suggest that the aggregate footprint of data centers on non-data center firms is limited. We note, however, that heterogeneous responses remain possible. Specialist IT providers may benefit from proximity to large data centers-for example, through lower latency or cheaper wholesale computing-whereas other establishments may face tighter competition for talent.

Taken together, our findings do not support the commonly held view that data center entry generates broad commercial gains for the host economy (Joint Legislative Audit and Review Commission, 2024). Over the medium horizon we study, establishments in the county neither expand nor contract materially in response to data center entry. [28]

We also extend the analysis by using commuting zones, as defined by Fowler (2024), as our geographical unit.[29] Commuting zones capture broader spillover effects, since they cover a larger territory than counties, though identification is weaker at this level. In particular, the instruments are not relevant, so we restrict attention to the reduced form.

[28] Due to the data provider change in 2020, we also estimate the model separately before and after 2020. Our null result with insignificant coefficients of interest prevails.

[29] The delineation is based on 2020 definitions. While some observations predate 2020, the alternative 2010 definitions are less suitable.

Table 12: Continuous DiD results of local economic outcomes with and without instrumenting

Number of observations: 8,864		**Method:** doubly-robust		**Polynomial order:** 1			
Outcome	Statistic	Estimate	SE	LB CI	UB CI	Switchers	Stayers
IT budget by revenue	AS	28.33337	20.29768	-11.45009	68.11683	939	7,925
	WAS	-0.0793772	0.1299997	-0.3341767	0.1754223	939	7,925
	IV-WAS	-0.485747	5.133717	-10.54783	9.576338	8,221	634
Employment by revenue	AS	0.0053822	0.0069677	-0.0082744	0.0190389	939	7,925
	WAS	-0.0000658	0.0000486	-0.0001609	0.0000294	939	7,925
	IV-WAS	0.0001394	0.0016311	-0.0030576	0.0033364	8,221	634
Log revenue	AS	-0.0003418	0.0006742	-0.0016632	0.0009796	939	7,925
	WAS	0.0000111	0.0000053	0.0000007	0.0000215	939	7,925
	IV-WAS	0.0000649	0.0002202	-0.0003667	0.0004965	8,221	634

Table 12 shows that all coefficients are statistically not different from zero, except for the weighted average slope of IT spending, which is estimated at 0.07. This reflects the average change in the outcome per one-unit increase in the continuous treatment, estimated only from periods in which some commuting zones (switchers) increase their data center load while others with the same initial load (stayers) do not.

Economically, this implies that firms in commuting zones experiencing additional data center demand allocate, on average, 7 cents more of every revenue dollar to IT spending than comparable zones without the increase. While this result is correlational, it is consistent with the expectation that firms located near large data centers invest more heavily in IT.

8 Impact on the environment

This section examines the local environmental impact of data centers by focusing on the CO_2e emissions of electricity plants at the county level. Specifically, we match data centers with power plants located in the same county. The outcome variable is county level CO_2e emissions

Table 13: Continuous DiD of SWZD outcomes based on commuting zones

Number of observations: 2,946		**Method:** doubly–obust		**Polynomial order:** 1			
Outcome	Statistic	Estimate	SE	LB CI	UB CI	Switchers	Stayers
IT budget by revenue	AS	78.74552	43.95249	-7.401354	164.8924	764	2,182
	WAS	0.0711342	0.0220083	0.0279978	0.1142705	764	2,182
Employment by revenue	AS	0.0024659	0.003154	-0.0037159	0.0086477	764	2,182
	WAS	2.36e-06	3.07e-06	-3.65e-06	8.37e-06	764	2,182
Log revenue	AS	-0.0002888	0.0006417	-0.0015466	0.0009689	764	2,182
	WAS	-3.36e-07	5.81e-07	-1.47e-06	8.02e-07	764	2,182

from electricity generation, while the treatment remains the electricity load of data centers in that county. The key question is whether data centers increase emissions strictly locally. This perspective likely to be most relevant for local decision-makers, who may disregard externalities spilling over to other counties. Table 14 summarizes the joint presence of data centers and power plants in 2020 and shows that counties with data centers almost always also host a power plant.

Table 14: Crosstabulation of `data center presence` and `power plant presence` in a county

	`power plant presence`		
`data center presence`	0	1	Total
0	953	1,813	2,766
1	20	290	310
Total	973	2,103	3,076

Table 15 reports the results using the same continuous DiD framework as in previous sections. The coefficient of interest is insignificant in both the baseline and IV specifications, and all models pass the placebo (pre-trend) test. These findings indicate that data centers do not increase local emissions. However, because electricity is not always sourced locally, emissions may rise elsewhere on the grid.

Table 15: Continuous DiD of CO2e emissions

Number of observations: 21,510	**Method:** doubly-robust	**Polynomial order:** 1					
Outcome	Statistic	Estimate	SE	LB CI	UB CI	Switchers	Stayers
total CO2e							
(baseline)	AS	130.487	301.0464	-459.5639	720.5379	1,407	20,103
	Placebo_1 AS	71.3044	44.14702	-15.2238	157.8326	51	17,115
	WAS	6.709762	5.66115	-4.3861	17.8056	1,407	20,103
	Placebo_1 WAS	-0.413153	0.424180	-1.2445	0.4182	51	17,115
total CO2e							
(with IV)	First stage WAS	13,069.07	4,723.695	3,810.633	22,327.52	18,421	3,069
	IV-WAS	-1.238914	22.56257	-45.4616	42.9837	18,421	3,069
total CO2e							
(IV + placebo							
& bootstrap)	First stage WAS	13,069.07	4,723.695	3,810.633	22,327.52	18,421	3,069
	Placebo_1 WAS	3,607.254	33,363.87	-61,785.92	69,000.43	1,931	1,105
	IV-WAS	-1.238914	10.26146	-21.5691	13.2062	18,421	3,069
	Placebo_1 IV-WAS	-39.83286	1,283.842	-2,556.164	2,476.498	1,886	1,015

Notes: Our baseline specification in the first block of the table shows that the DC energy load does not significantly increase the CO2e emissions in the same county. Our IV specification further strengthens this point, with the second block showing the estimation with the theoretical standard errors, while the third block is with bootstrapped standard errors. All our specifications pass the placebo test.

9 A Small Regional Model of AI & DC Demand

We use a deliberately small partial-equilibrium model of one balancing area (RTO/ISO zone or large utility territory) to map exogenous growth in data-center (DC) load into wholesale prices, regulated retail tariffs, and emissions. The wholesale side clears on an upward-sloping supply curve with scarcity when peak load approaches derated capacity. A regulated distributor then recovers energy and non-energy costs via class-specific tariffs.

Two model primitives govern retail outcomes: (i) a wholesale pass-through parameter λ_k that captures how much of Δp^W flows into class-k rates between rate cases, and (ii) an allocation share α_k that determines how new non-energy revenue requirements (capacity, T&D, interconnection) are spread across classes. A third primitive, the DC procurement share β, splits DC load between private PPAs and the grid. Calibrated to our pre-2023 results, the model nests the historical null (low λ and modest α) and then evaluates forward-looking scenarios under

accelerated AI/DC demand.

Environment and Loads There are three customer classes: households H, non-DC commercial/industrial firms F, and data centers D. A regulated utility procures energy and capacity from the wholesale market and recovers an incremental revenue requirement through class-specific tariffs. Let L_t^k denote annual energy (MWh) for class $k \in \{H, F, D\}$. A share $\beta \in [0, 1]$ of DC load is covered by off-system PPAs or behind-the-meter supply; the remainder uses grid supply.

Total load that the RTO must serve is

$$Q_t \;=\; L_t^H \;+\; L_t^F \;+\; (1-\beta)\, L_t^D. \tag{4}$$

Let Q_t^{peak} be the annual peak (MW), e.g., $Q_t^{\text{peak}} = \phi\, Q_t/8760$ with peak factor $\phi > 1$.

Wholesale Price and Capacity Wholesale price p_t^W is determined by a reduced-form upward-sloping supply curve with scarcity at tight capacity:

$$p_t^W \;=\; \begin{cases} a + b \cdot Q_t^{\text{peak}}, & \text{if } Q_t^{\text{peak}} \le \kappa K_t, \\ p^{\text{scar}}, & \text{if } Q_t^{\text{peak}} > \kappa K_t, \end{cases} \tag{5}$$

where a and b are the intercept and slope of the merit-order, K_t is nameplate capacity (MW), $\kappa \in (0, 1]$ is an effective derating factor, and p^{scar} is a scarcity price cap.

In the long run, capacity expands to satisfy a reserve margin RM (MW). A simple entry rule is

$$K_{t+1} \;=\; \max\left\{K_t,\; \frac{Q_{t+1}^{\text{peak}} - \text{RM}}{\kappa}\right\}. \tag{6}$$

Regulated Retail Tariffs and Incidence Let the utility's annual incremental non-energy revenue requirement be

$$\Delta RR_t^{\mathrm{NE}} \;=\; \underbrace{\mathcal{A}(r,N)\left(1-\varphi\right)c_K\,\Delta K_t}_{\text{annualized capex added to rate base}} \;+\; \underbrace{c_K^{\mathrm{OM}}\,\Delta K_t}_{\text{fixed O\&M}} \;-\; \Delta\mathrm{OtherRev}_t, \tag{7}$$

where $\Delta K_t \equiv K_t - K_{t-1}$, c_K is capex per MW, $\varphi \in [0,1]$ is the DC contribution (contributions in aid of construction), $\mathcal{A}(r,N)$ is the annuity factor at allowed return r over life N, and c_K^{OM} is fixed O&M per MW-year.

Retail price for class $k \in \{H,F\}$ adjusts via two primitives: (i) wholesale pass-through $\lambda_k \in [0,1]$, and (ii) allocation share α_k of the non-energy requirement, with $\sum_{k\in\{H,F\}} \alpha_k = 1$ (DC contract recovery is netted in $\Delta RR_t^{\mathrm{NE}}$). Linearizing the change in the class-k average tariff (in \$ per MWh),

$$\Delta p_{k,t} \;\approx\; \lambda_k\,\Delta p_t^W \;+\; \frac{\alpha_k\,\Delta RR_t^{\mathrm{NE}}}{S_{k,t-1}}, \qquad S_{k,t-1} \equiv L_{t-1}^k \;\;(\mathrm{MWh}), \tag{8}$$

where $\Delta p_t^W \equiv p_t^W - p_{t-1}^W$.

Short vs. long run. In the short run (no new K), $\Delta RR_t^{\mathrm{NE}} = 0$ and only wholesale pass-through operates. In the long run, capacity catches up via (6), generating $\Delta RR_t^{\mathrm{NE}} > 0$ and possibly lowering p^W relative to the short run.

Firms and Emissions For non-DC firms, let electricity have cost share s_E and short-run demand elasticity ϵ_E. A convenient mapping is

$$\frac{\Delta Y_t^F}{Y_{t-1}^F} \;\approx\; -\,s_E\,\epsilon_E \cdot \frac{\Delta p_{F,t}}{p_{F,t-1}}. \tag{9}$$

Emissions derive from dispatch shares $s_{i,t}$ across technologies i with emissions factor ef_i:

$$\Delta \mathrm{CO_2e}_t \;=\; \gamma_t \,(Q_t - Q_{t-1}) \;+\; \sum_i ef_i \,(G_{i,t} - G_{i,t-1}), \qquad \gamma_t \equiv \sum_i s_{i,t}\, ef_i. \tag{10}$$

Shock Process and Solution Data-center load follows an exogenous path, e.g.,

$$L_t^D \;=\; L^{D,0} \cdot (1+g_D)^t \quad \text{or} \quad L_{t+1}^D \;=\; \xi\, L^{D,0}, \;\; \xi \in \{3, 10\}. \tag{11}$$

Each period: (i) observe L_t^D, compute Q_t and Q_t^{peak}; (ii) set p_t^W via (5); (iii) update tariffs via (8); (iv) if needed, update capacity via (6). The historical "no-impact" case nests as $\lambda_k \approx 0$ and/or small α_k.

Calibration The model is disciplined to replicate our pre-2023 evidence that data-center load did not move local retail tariffs, firms' outcomes, or county CO_2e. Concretely, we choose the wholesale pass-through parameters λ_H and λ_F so that, when we feed the model the observed 2017-2023 path of county data-center load, the simulated changes in residential and commercial prices are statistically indistinguishable from zero, matching the continuous-treatment DiD and event-study results in the previous sections (Tables 2-4, 8-11; Figures 5 and 7).

We pin the cost-allocation shares α_k and the non-energy revenue requirement to be negligible in-sample by calibrating reserve margins and capacity entry so that the realized peak increments do not trigger rate-base additions attributable to data centers-consistent with the flat retail incidence we document. We set the procurement share β to reflect the institutional evidence that a material fraction of hyperscale demand is met via PPAs/nuclear contracts rather than entirely through the local utility, which helps reconcile zero county-level price effects with growing compute demand (Section 2).

The supply-curve slope (a, b) is recovered from pre-2023 wholesale price-load data so the wholesale side reacts plausibly to load, while the firm block uses near-zero short-run elasticities

to match the null effects on ICT spending, revenue, and employment (Section 7, Table 12).

Finally, the emissions module is benchmarked to RTO-level marginal emissions so that county-level emissions are insensitive to local data-center load, reproducing the null in Section 8 (Table 15). Under this parameterization the model exactly nests the historical null: when we replay the realized 2017-2023 load shocks, the simulated paths for prices, firms, and local emissions are flat, and departures from these calibrated primitives generate the forward-looking scenarios we study next.

Simulations Table 16 presents our simulation results in case of shocks to energy demand for 5 different scenarios with high and low pass-through to residential energy prices.

Three messages emerge. First, the household impact is small when most DC energy is procured via PPAs or when pass-through is muted. In those cases a large increase in DC load leaves retail bills nearly unchanged; the grid sees the energy, but rate design insulates households.

Second, policy dials matter more than raw load: the same tenfold demand shock becomes material for households only when it is *grid-sourced* (low β) and when regulators allow *generous pass-through* and allocate a visible share of new non-energy costs to the residential class (higher λ and α). Under that configuration, household bills move from no change to noticeable monthly additions.

Third, even when wholesale prices look similar in the short and long run, the long-run bill can exceed the short-run bill because new network and interconnection investments enter the rate base. In our calibration this long-run wedge is modest when DCs finance a large share of upgrades, and larger when networks are expanded with limited direct DC contributions. Together these patterns reconcile our historical null with forward-looking concern: institutions, not just megawatts, determine who pays as AI demand scales.

Table 16: Impact on wholesale and residential energy prices of shocks to DC energy demand

Main scenario results (prices in ¢/kWh, bills in $/month):

Scenario	β	λ_H	α_H	p_{W0}	p_{W1}^{SR}	p_{W1}^{LR}	$\Delta p_{c/MWh}^{SR}$	$\Delta p_{c/MWh}^{LR}$	$\Delta Bill_{mo}^{SR}$	$\Delta Bill_{mo}^{LR}$	RR_{NE}^{M}
"10x grid-sourced"	0.1	0.2	0.2	380.75	494.45	494.45	2.274	2.615	20.47	23.53	852.09
"10x, 70% PPA"	0.7	0.1	0.1	380.75	412.26	412.26	0.315	0.362	2.84	3.26	236.12
"10x, high pass-thr"	0.1	0.6	0.4	380.75	494.45	494.45	6.822	7.504	61.40	67.53	852.09
"3x, renew+storage"	0.6	0.2	0.2	380.75	387.60	387.60	0.137	0.158	1.23	1.42	51.33
"3x, status quo"	0.3	0.1	0.2	380.75	399.93	399.93	0.192	0.249	1.73	2.24	143.73

10 Policy implications and conclusions

Our results suggest that, despite their large commercial load, data centers impose minimal spillovers on local retail electricity prices and only modest effects on non-data-center firms. Regulators and utility planners can take confidence that existing rate-setting frameworks and grid planning processes have so far absorbed new data center demand without driving up household or commercial rates in any meaningful way. That said, as data center growth continues, grid operators should monitor local capacity constraints and consider targeted investments-such as substation upgrades or demand response programs-to maintain reliability and keep wholesale costs in check.

From an economic development perspective, the absence of adverse impacts on firms' ICT spending, revenue, or overall employment (beyond a small shift of IT workers into data centers) indicates that hosting data centers did not crowd out other business activity, at least up to 2023.

However, the future may look very different from the past. As the public debate now focuses on whether accelerating AI demand will raise local bills, we build a simple model that speaks directly to that question and reframes it as a regulatory problem. The simulations show that spillovers to the local economy hinge on these rules: with PPA-heavy procurement and muted pass-through, household bills remain effectively insulated even under large shocks; when growth is grid-sourced and regulators both transmit wholesale changes and spread T&D/interconnection costs to the residential class, bills become noticeable. Any long-run wedge

comes from how infrastructure enters the rate base, not from inexorable wholesale price pressure.

In summary, data centers appear to source most of their power from the commercial grid at passthrough rates close to unity, but do not translate into higher retail rates or reduced firm performance, at least up to 2023. Simulations show that this could change under regulatory regimes that permit high wholesale pass-through and allocate network upgrades to households when demand is grid-sourced, making incidence a policy choice rather than a technological inevitability. Future work could examine long run grid investments and distribution level effects, as well as environmental outcomes under different power procurement strategies.

References

H. Averch and L. L. Johnson. Behavior of the Firm Under Regulatory Constraint. Technical Report 5, 1962.

K. Borusyak, X. Jaravel, and J. Spiess. Revisiting Event-Study Designs: Robust and Efficient Estimation. *Review of Economic Studies*, 11 2024. ISSN 0034-6527. doi: 10.1093/restud/rdae007.

B. Callaway and P. H. Sant'Anna. Difference-in-Differences with multiple time periods. *Journal of Econometrics*, 225(2):200–230, 12 2021. ISSN 18726895. doi: 10.1016/j.jeconom.2020.12.001.

C. De Chaisemartin and X. D'haultfoeuille. Difference-in-Differences Estimators of Intertemporal Treatment Effects *. 2020.

C. de Chaisemartin and X. D'Haultfœuille. Two-Way Fixed Effects Estimators with Heterogeneous Treatment Effects. *American Economic Review*, 110(9):2964–2996, 9 2020. ISSN 19447981. doi: 10.1257/aer.20181169.

C. De Chaisemartin, X. D'haultfoeuille, F. Pasquier, D. Sow, G. Vazquez-Bare, and M. Farrell. Difference-in-Differences for Continuous Treatments and Instruments with Stayers *. 2025.

N. Fabra and M. Reguant. Pass-Through of Emissions Costs in Electricity Markets. Technical report.

C. S. Fowler. New Commuting Zone delineation for the U.S. based on 2020 data. *Scientific Data*, 11(1), 12 2024. ISSN 20524463. doi: 10.1038/s41597-024-03829-5.

S. Greenstein and T. P. Fang. Where the Cloud Rests: The Economic Geography of Data Centers. Technical report, 2020.

Z. He, S. Jiang, D. Xu, and X. Yin. Investing in Lending Technology: IT Spending in Banking *. Technical report, 2023.

Husch Blackwell. Tax Incentives for Data Centers 50 State Survey Contents. Technical report, 2025.

Joint Legislative Audit and Review Commission. Data Centers in Virginia. Technical report, 2024.

E. Martin and A. Peskoe. Extracting Profits from the Public: How Utility Ratepayers Are Paying for Big Tech's Power. Technical report, 2025.

A Data sources

A.1 Aberdeen Technology Data Cloud

Our primary data source for establishment-level information is the Aberdeen Technology Data Cloud database, previously known as the Harte Hanks, CI Technology Database (CiTDB), and Spiceworks Ziff Davis (Aberdeen) database. This database was initiated in the 1980s and has undergone several transitions in its methodology over the years.

Until 2015, trained research assistants conducted over 50,000 interviews each month to ensure the accuracy and reliability of the data. In addition to direct responses, the data providers inferred information across 20 technology areas, ranging from total installed PCs to the likely presence of WAN infrastructure. Covering businesses from single-location firms to Fortune 1000 enterprises, the database focuses on identifying business locations, or "sites," that have potential future IT purchases, along with the decision-makers who influence and authorize major technology investments.

Respondents to CiTDB's syndicated interviews are IT professionals with firsthand knowledge of the technology present at their locations. However, from 2016 onward, the database has shifted its emphasis from direct interviews toward a greater reliance on inferred data, leveraging advanced analytical techniques to estimate values more efficiently.

Our main variable of interest, which is the IT budget of each site, consist of the following four main components: Software budget, hardware budget, communication budget, service budget.

The definition of these budgets are as follows based on He et al. (2023).

A.2 Communication budget

The modeled IT budget for communication services at this site. It is defined as the network equipment that companies operate to support their communications needs, including routers, other line equipment, fiber optic equipment, switches, private branch exchanges (PBXs), radio

and TV transmitters, fax machines, desktop telephone sets, wide area network (WAN) and local-area network (LAN) equipment, videoconferencing and telepresence equipment, radio equipment other network equipment, and client-mobile devices like cell phones/Pherios that are bought by individuals.

The modeled IT budget for software at this site. It is defined as software (not 'Web goods,' whether that software is packaged or unpackaged) software revenue (bought directly) rather than remotely hosted by a third party, offered for a flat-rate from a multitenent shared instance where accessed by a browser, or customized by a company by themselves (customized systems), including license, maintenance, subscription, and software vendor-provided services revenues for all software types, such as: operating systems, middleware, development, security and management systems, IT management systems, security software, application servers and business intelligence, application software such as desktop applications, business administrative software (like business intelligence and enterprise content management), vertical applications (like ERP, CRM, SCM or PLM), eLearning software, collaboration software, vertical industry applications (like banking management systems, security trading systems, insurance underwriting or claims management software, retail management systems, hospital management systems), even though their cost is often bundled, and computer operating systems software (even though that cost is often bundled). The modeled IT budget for ITservices services at this site.

It is defined as project-based consulting or systems integration services that vendors provide to businesses and Governments, whether on or off-site, including contractors, consulting services for IT strategy, security assessments and process change, systems integration, applications outsourcing, programs outsourcing, desktop support outsourcing, distributed systems outsourcing, network management, systems architecture, application management outsourcing and application testing, application site implementations software deployments, generally used rather than subscribed to, and thus are different from SaaS, and computer hardware support and maintenance services. The modeled IT budget for hardware at this site. It includes the

(large) computer hardware that IT departments buy and support, regardless of whether the IT department itself operates that equipment (can be in-service or oversees the use of this equipment by employees built-in PCs), including PCs (personal computers (laptops, desktops, and tablets)), server hardware, peripherals: monitors, terminals, printers, keyboards, mice, USB devices, etc., storage: storage devices (NAS, DAS, tape), and other hardware: this refers to industry specific IT hardware (point-of-sales equipment based on PCs, smart cards, embedded computer chips, etc.).