Do Data Centers Raise Residential Electricity Prices? Evidence from 24 U.S. States

Kiran Garimella

Preliminary Draft: June 2026

Abstract

Data centers are among the fastest-growing sources of electricity demand in the United States, yet their effect on residential electricity prices is not well established. We build a state-ZIP-level panel of 22,834 units across 24 states for 2014–2024, linking 2,277 analytical-sample data centers scraped from interconnection queue filings to household electricity expenditure data from the American Community Survey Public Use Microdata Sample (ACS PUMS). We apply four causal inference strategies (two-way fixed effects difference-in-differences, event study, matched DiD, and dose-response estimation), a robustness analysis at the utility service territory level (the unit at which residential rates are actually set), and a battery of further checks: megawatt-based treatment intensity, time-varying household income controls, subsamples by market structure, hyperscaler heterogeneity, a placebo test with random treatment assignment, a utility-territory spillover test, and a cohort DiD ATT diagnostic. The baseline ZIP-level TWFE estimate is a \$2.09 per month increase in electricity bills associated with data center presence, marginally significant ($p = 0.083$). The matched DiD on comparable state-ZIPs yields an insignificant \$0.90. The ZIP-level event study shows delayed positive coefficients at four and five years after entry, but re-estimating at the utility territory level (602 territories) produces insignificant TWFE effects, including a household-weighted estimate of $-\$2.00$ ($p = 0.605$).

The additional checks support a cautious interpretation: the baseline ZIP-level estimate falls to \$1.84 ($p = 0.122$) once time-varying income is controlled, becomes negative and significant when treatment is measured in megawatts ($-\$0.011$ per MW, $p = 0.001$), is insignificant in every market-structure subsample, and is not supported by a same-utility spillover test. Overall, the evidence does not robustly support the claim that data centers materially increased residential electricity bills during the pre-hyperscaler period.

Two limitations bound how far these results travel. The data center inventory, drawn from interconnection queue filings, likely captures only a subset of operational facilities and biases estimates toward zero. The panel also ends in 2024, before the hyperscale facilities (500 MW to 1 GW+) planned for 2025 onward come online. Null findings from this pre-hyperscaler era should not be extrapolated to the larger facilities now under construction. We discuss how regulated rate structures, grid-level pricing mechanisms, and the geographic concentration of data centers may attenuate local price effects, and consider implications for ongoing policy debates about data center siting and electricity infrastructure planning.

Contents

1 Introduction

Cloud computing, artificial intelligence, and digital services have driven a sharp increase in data center construction across the United States. Data centers, the physical infrastructure housing the servers, storage systems, and networking equipment that support the digital economy, now consume more electricity than many small countries. The International Energy Agency estimates that global data center electricity consumption reached approximately 460 terawatt-hours (TWh) in 2024, or roughly 2% of global electricity demand, and projects that the figure could more than double by 2030 (International Energy Agency, 2024). In the United States alone, data centers consumed an estimated 176 TWh in 2023, approximately 4.4% of total national electricity consumption, with projections suggesting that the share could reach 6 to 12% by 2028 (Shehabi et al., 2024).

This growth has produced public concern about residential electricity prices. As data centers compete with households for grid capacity and generation, whether their concentrated demand drives up costs for residential ratepayers has become a central issue in energy policy debates. In Virginia, home to the world's largest concentration of data centers in Loudoun County's "Data Center Alley," Dominion Energy requested rate increases that would raise the typical residential bill by approximately 10% in 2024, prompting public scrutiny of whether data center growth was a contributing factor (Paullin, 2023). Similar concerns have surfaced in Oregon, where data center growth in Hillsboro and Prineville has strained local grid capacity, and in Texas, where the Electric Reliability Council of Texas (ERCOT) has had to integrate hundreds of new large-load interconnection requests (Rogoway, 2024; Blunt and Hiller, 2026).

Despite the intensity of public debate, empirical evidence on the causal relationship between data center presence and residential electricity prices is thin. Most existing analyses rely on aggregate trends, anecdotal evidence, or simple cross-sectional correlations that cannot distinguish causation from selection. Data centers locate in areas with low electricity costs, favorable tax environments, and capable grid infrastructure, all of which create an identification challenge. To our knowledge, no prior study has attempted a systematic causal analysis exploiting the staggered timing of data center entry across

ZIP codes to identify the effect on household electricity expenditures.

This paper fills that gap. We construct a state-ZIP-level panel dataset of 22,834 units across 24 U.S. states for 2014–2024 (excluding 2020, when the ACS data collection was disrupted by the COVID-19 pandemic). The panel links two primary sources: (1) information on 2,277 analytical-sample data centers (location, capacity, and operational dates) scraped from interconnection queue filings compiled by Interconnection.fyi, and (2) household-level electricity expenditure data from the American Community Survey Public Use Microdata Sample (ACS PUMS), aggregated to the ZIP code level using Census Bureau geographic crosswalk files.

Our identification strategy exploits the staggered timing of data center entry. We use four causal inference approaches at the ZIP code level and one robustness analysis at the utility service territory level:

1. **Two-Way Fixed Effects Difference-in-Differences (TWFE DiD):** Our baseline specification includes ZIP code and year fixed effects, absorbing time-invariant ZIP characteristics (geography, demographics, housing stock, state-level baseline rates) and nationwide time trends (inflation, energy market conditions). We estimate both the extensive margin (any data center presence) and the intensive margin (log cumulative data center count).
2. **Event Study Analysis:** We estimate dynamic treatment effects for "clean treated" state-ZIPs, those receiving their first data center during the panel period, relative to never-treated controls. The event study serves as a diagnostic for the parallel trends assumption and reveals the temporal pattern of any treatment effects.
3. **Matched Difference-in-Differences:** To address selection on observables, we match each treated state-ZIP to its five nearest neighbors among control state-ZIPs based on pre-treatment electricity expenditures, then re-estimate the DiD on the balanced subsample.
4. **Dose-Response Analysis:** Among "always-treated" state-ZIPs that had data centers before our panel begins, we use variation in the rate of additional data center construction to test whether the intensive margin of data center accumulation

predicts bill increases.

5. **Utility Territory-Level Robustness Analysis:** Because electricity rates are set at the utility service territory level rather than the ZIP code level, we re-estimate our core specifications using 602 utility territories as the unit of analysis. The territory-level specification addresses a potential SUTVA violation that affects the ZIP-level approach.

The main findings are as follows. The baseline TWFE estimate associates data center presence with a \$2.09 per month increase in residential electricity bills, but the effect is only marginally significant at the 10% level ($p = 0.083$) and is not significant at conventional levels. The corrected event study does not show statistically significant pre-treatment divergence, but it does show delayed positive coefficients four and five years after data center entry. The matched DiD, which compares treated state-ZIPs to observably similar controls, yields a smaller and statistically insignificant estimate of \$0.90 per month ($p = 0.464$). The dose-response analysis among always-treated state-ZIPs finds no evidence that additional data center construction raises bills; the log cumulative data center coefficient is significantly negative, which we attribute to mean reversion and compositional effects rather than a genuine protective effect of data center density.

We also re-estimate at the utility service territory level, the geographic unit at which residential electricity rates are actually determined. We consider this the more institutionally appropriate specification, since it aligns the unit of analysis with the level at which rates are set and avoids the within-territory SUTVA violations that affect the ZIP-level approach. The utility-level TWFE finds no significant effect in either the unweighted specification (\$0.35, $p = 0.834$) or the household-weighted specification ($-\$2.00$, $p = 0.605$). The utility-level event study shows no statistically significant pre-treatment divergence, but the post-treatment estimates are imprecise and do not show a monotonic build-up. Consistent weak-to-null TWFE results at both levels of aggregation strengthen the main conclusion, though confidence intervals are wide enough that modest positive effects cannot be ruled out.

To probe these conclusions further, we run seven additional robustness checks (Section 5.4). The baseline ZIP-level estimate falls to \$1.84 ($p = 0.122$) once time-varying household income is included as a control, and becomes negative and significant when treatment is measured by installed megawatts rather than by facility count (−\$0.011/MW, $p = 0.001$). The presence coefficient is insignificant in every market-structure subsample (regulated, restructured, hybrid), is not increased by separating out hyperscaler facilities, and is accompanied by a cohort DiD ATT diagnostic of \$2.44. A placebo test that randomly assigns treatment year to never-treated state-ZIPs returns a coefficient distribution centred on zero with a standard deviation of \$0.99. A same-utility spillover test does not support the earlier same-3-digit-ZIP spillover interpretation: never-treated state-ZIPs in utilities with clean-treated data-center entry have a negative spillover coefficient (−\$1.37, $p < 0.001$).

These null-to-weak findings may seem at odds with public discourse, but they are consistent with several features of U.S. electricity markets. Most states in our sample operate under cost-of-service regulation, in which rates are set by public utility commissions through formal rate cases that spread infrastructure costs across broad customer classes, rather than reflecting localized demand. Data centers also typically connect to high-voltage transmission networks and are served under separate large-commercial or industrial tariff schedules, with limited direct cost allocation to residential ratepayers.

We stress two caveats. First, our inventory of 2,277 analytical-sample data centers, sourced from interconnection queue filings, almost certainly undercounts the true population of U.S. data centers. Many older, smaller, or privately operated facilities do not appear in public filings. The undercount biases our estimates toward zero and means that some control ZIP codes may contain unobserved data centers. Second, our panel ends in 2024, before the wave of large new construction comes online. The hyperscale facilities of 500 MW to 1 GW+ planned by major cloud and AI companies for 2025 onward are an order of magnitude larger than the typical data centers in our sample. Our weak-to-null findings reflect the era of moderate data center growth and should not be projected onto the hyperscaler era, whose grid-level cost impacts may take years to materialize in residential rates.

This paper contributes to several literatures. It is, to our knowledge, the first large-scale causal analysis of data center effects on residential electricity expenditures that uses geographic and temporal variation across ZIP codes and utility territories. It also shows that even very large industrial loads may have limited pass-through to residential rates under regulated pricing. And it offers evidence on a policy claim, that data centers "raise electricity bills," that is widely repeated but not well tested at the local level, with the caveats about grid-level and long-run effects noted above.

The remainder of the paper is organized as follows. Section 2 reviews the literature on data center energy consumption, electricity pricing, and causal inference in energy economics. Section 3 describes the data sources, collection methodology, and the panel construction. Section 4 presents the empirical strategy. Section 5 reports the findings across all four identification strategies. Section 7 discusses implications, limitations, and directions for future research. Section 8 concludes.

2 Related Work

This paper sits at the intersection of several literatures: data center energy consumption and sustainability, electricity pricing and rate design, the local economic impacts of large industrial facilities, and causal inference methods in energy economics. We review each in turn.

2.1 Data Center Energy Consumption

The energy footprint of data centers has been the subject of growing academic and policy attention. Early work by Koomey (2011) estimated that global data center electricity consumption roughly doubled between 2000 and 2005, raising alarm about exponential growth trajectories. Later studies moderated these projections by documenting efficiency improvements. Masanet et al. (2020) showed that despite a six-fold increase in data center compute output between 2010 and 2018, total energy consumption grew by only 6%, owing to the shift toward hyperscale facilities, improvements in server efficiency (including the decline of "comatose" servers), and advances in cooling technology. Shehabi et al. (2016) at Lawrence Berkeley National Laboratory (LBNL) developed bottom-up models of U.S. data center energy use, estimating consumption at approximately 70 billion kWh in 2014, or about 1.8% of total U.S. electricity consumption.

More recent assessments have revised these figures upward. LBNL's 2024 update (Shehabi et al., 2024) estimated U.S. data center energy consumption at 176 TWh in 2023, an increase driven by the rapid build-out of AI training and inference infrastructure. The International Energy Agency's *Electricity 2024* report (International Energy Agency, 2024) projected that global data center electricity demand could reach 1,000 TWh by 2026 under aggressive growth scenarios, equivalent to Japan's total electricity consumption. de Vries (2023) raised concerns about the energy intensity of generative AI workloads, estimating that a single ChatGPT query could consume 3 to 10 times more electricity than a conventional Google search.

These studies document aggregate energy consumption but generally do not address the downstream effects on electricity prices. We bridge that gap by directly testing whether the local presence and accumulation of data centers affects household electricity expenditures.

2.2 Electricity Pricing and Rate Regulation

Understanding why data centers might or might not affect residential electricity prices requires understanding the institutional structure of electricity rate-setting. In the United States, approximately two-thirds of electricity is sold by investor-owned utilities (IOUs) subject to cost-of-service regulation by state public utility commissions (PUCs) (Joskow, 2008b). Under this framework, utilities propose revenue requirements based on capital investments and operating costs, and PUCs set rates to allow a reasonable rate of return. Rates are typically differentiated by customer class (residential, commercial, industrial) but are uniform within a class across a utility's service territory, so localized demand conditions in a particular ZIP code do not directly affect prices.

Borenstein (2012) reviewed U.S. electricity pricing and showed that residential rates reflect a mix of generation costs, transmission and distribution charges, renewable energy mandates, and cross-subsidies between customer classes. Bushnell et al. (2017) showed that capacity markets and resource adequacy requirements create additional channels through which demand growth can affect long-run prices, as utilities must invest in new generation and transmission capacity to serve growing load.

Several states in our sample, including Texas, operate restructured (deregulated) markets where wholesale electricity prices are set competitively, and retail prices may more directly reflect local supply-demand conditions (Joskow, 2008a). Borenstein and Bushnell (2015) analyzed the Texas ERCOT market and showed that wholesale price variation, while significant, is substantially smoothed in retail rates. The distinction between regulated and restructured states is potentially important for our analysis, since data center demand effects may be more detectable in competitive markets than in regulated ones.

A policy-oriented literature has examined how large industrial loads (and data centers specifically) interact with rate structures. Electric Power Research Institute (2024) documented that many utilities have created special tariff schedules for data centers and other large loads, with demand charges, interruptibility provisions, and time-of-use pricing that differ from standard commercial rates. Several state PUCs have opened proceedings to examine whether data center growth necessitates changes to rate allocation methodologies, including Virginia's State Corporation Commission (Virginia State Corporation Commission, 2024), Georgia's Public Service Commission (Georgia Public Service Commission, 2026), and the Oregon PUC (Oregon Public Utility Commission, 2026).

2.3 Local Economic Impacts of Data Centers

A parallel literature examines the broader local economic impacts of data center development. Data centers are unusual industrial facilities: they involve very large capital investments ($500 million to several billion dollars for hyperscale campuses) but create relatively few permanent jobs (typically 30 to 100 per facility), generate significant tax revenue through personal property taxes on computing equipment, and impose environmental externalities through electricity consumption, water use for cooling, and noise (Virginia Joint Legislative Audit and Review Commission, 2024).

Miller (2021) studied the tax incentive programs that states and localities use to attract data centers and found that the net fiscal impact depends on the structure of incentives. Some jurisdictions effectively give away more in tax abatements than they receive in new revenue. Loudoun County Department of Economic Development (2023) documented that Loudoun County, Virginia (home to approximately 70% of the world's internet traffic) collected over $600 million in annual data center tax revenue by 2023, making it one of the wealthiest counties in the United States, while simultaneously experiencing rapid residential growth and infrastructure strain.

Community opposition to data center development has grown in recent years, with concerns about noise, water consumption, visual impact, and electricity costs. Citizen groups and consumer advocates in Virginia, Oregon, and Georgia have argued that data center electricity demand drives up rates for residential customers, but these claims have generally relied on temporal correlation rather than causal analysis (Paullin, 2023; Rogoway, 2024; Baker, 2024).

2.4 Causal Inference in Energy Economics

Our empirical approach builds on a methodological literature. The two-way fixed effects (TWFE) difference-in-differences estimator has been widely used in energy and environmental economics to evaluate the effects of policies and shocks. Callaway and Sant'Anna (2021) and de Chaisemartin and D'Haultfœuille (2020) showed that in settings with staggered treatment adoption and heterogeneous treatment effects, the standard TWFE estimator can produce misleading estimates due to "negative weighting" of some treatment effects. Sun and Abraham (2021) demonstrated that standard dynamic TWFE event study specifications are *also* subject to contamination from heterogeneous treatment effects across cohorts, and proposed an interaction-weighted (IW) estimator that addresses the issue by estimating cohort-specific average treatment effects on the treated

(CATTs) and aggregating them with appropriate weights. Our setting partially mitigates these concerns: treatment (data center entry) is highly persistent, and our primary identification comes from comparing clean treated units against never-treated controls. We acknowledge, however, that our standard TWFE event study does not fully resolve the negative weighting problem.

Event study designs have been used extensively in energy economics. Davis (2011) used an event study framework to estimate the effect of power plant openings on local housing prices and found significant negative effects that dissipate with distance. Jarvis (2022) used a similar approach to study the impact of wind farm installations on local electricity prices in Great Britain. Our event study uses never-treated units as controls and omits the $k = -1$ period as the reference category, following common practice. The standard dynamic TWFE specification is, as noted, subject to the contamination issues identified by Sun and Abraham (2021); we report a cohort-size-weighted group-time ATT diagnostic, using never-treated units as the comparison group, as an additional check.

Matching estimators, as used in our matched DiD specification, have been used to improve covariate balance in energy economics studies. Ho et al. (2007) discussed the rationale for combining matching with regression adjustment, arguing that preprocessing via matching can reduce model dependence and improve the robustness of causal estimates. Imbens and Rubin (2015) provided a longer treatment of matching methods in observational studies.

2.5 Data Centers and Electricity Prices: Direct Evidence

To our knowledge, there is very limited prior work directly estimating the causal effect of data centers on residential electricity prices. The U.S. Department of Energy (U.S. Department of Energy, 2024) identified AI-driven load growth as a grid-planning problem requiring better forecasting, faster capacity studies, and accelerated transmission and permitting analysis. Boston Consulting Group analysts (Lee et al., 2023) projected that data centers' share of U.S. electricity demand could triple by 2030 as generative AI workloads grow, but that projection was based on aggregate supply-demand modeling rather than empirical estimation from observed outcomes.

News investigations have explored the relationship qualitatively. Reporting by Blunt and Hiller (2026) documented that utilities in Virginia, Georgia, and the Carolinas were requesting large rate increases partly attributed to data center-driven grid expansion. Penn and Weise (2025) profiled community opposition in rural Oregon, where residents reported that local grid congestion from data center operations coincided with more frequent brownouts and higher bills.

Our paper contributes to this thin literature by providing the first systematic, multi-state causal analysis exploiting the staggered timing of data center entry across ZIP codes, a well-defined identification strategy, and multiple robustness checks including analysis at the utility territory level.

3 Data

The panel combines three sources: data center locations and operational dates from Interconnection.fyi, household electricity expenditures from the American Community Survey (ACS) Public Use Microdata Sample (PUMS), and PUMA-to-ZIP and ZIP-to-utility crosswalks from public Census and EIA mappings. Appendix A lists URLs and variable names.

3.1 Data Centers

We scrape Interconnection.fyi for the 24 states in the sample.[1] Interconnection.fyi aggregates interconnection-queue filings from utilities and regional grid operators and is, as far as we know, the most complete public source for data center locations, capacities, and operational dates. We restrict to facilities with "Operational" status, valid in-state ZIP codes, and operational dates no later than 2024, leaving **2,277 analytical-sample data centers** across 23 states (Connecticut has none). Table 1 reports counts by state. The

[1]The 24 states are: Arizona, California, Colorado, Connecticut, Florida, Georgia, Illinois, Indiana, Iowa, Michigan, Minnesota, Missouri, Nevada, New Jersey, New York, North Carolina, Ohio, Oregon, Pennsylvania, Tennessee, Texas, Virginia, Washington, and Wisconsin. They include every major U.S. data center market and span the range of regulatory regimes and market structures.

distribution is heavily skewed: Texas (651), Virginia (552), and California (438) together hold over 72% of the sample, while 11 states have fewer than 15 facilities each.

Table 1: Data Centers by State

State	DCs	ZIP Codes	State	DCs	ZIP Codes
TX	651	1,995	MN	10	896
VA	552	915	IA	10	972
CA	438	1,808	NC	8	853
OR	268	430	TN	7	638
AZ	72	417	FL	6	1,013
IL	51	1,399	IN	5	808
NJ	50	599	MI	5	994
CO	33	530	NV	5	182
GA	27	756	WI	4	787
NY	18	1,819	WA	15	604
OH	18	1,237	CT	0	289
MO	13	1,045	PA	11	1,848
Total: 2,277			**Total state-ZIPs: 22,834**		

Note: Operational data centers listed on Interconnection.fyi as of March 2026, restricted to valid ZIP-coded facilities operational by 2024. ZIP counts are state-ZIP units with ACS PUMS bill data; border ZIPs appearing in multiple states are retained as distinct state-ZIP units.

For each facility we observe the facility name, location (city, county, state, often street address), capacity in megawatts (in binned form), permit status, grid operator, and a sequence of dates: application, approval, proposed operational, and actual operational. Each record also carries a flag for hyperscaler status (Yes/Likely/No/Unknown) that we use in the heterogeneity analysis of Section 5.4.

Geographic assignment. About a third of records lack a ZIP code in the listing. We geocode the address and assign the resulting coordinates to a ZIP using a public geocoding service, with manual review of any ZIP that crosses a state boundary. Records that fail geocoding are excluded from the ZIP-level analysis but kept in state counts.

Operational year. We use the actual operational date when available; otherwise the approval date, then the application date. Facilities with no dates but a permit status of "Existing" are assigned an operational year of 2013 (pre-panel), as are the remaining records with no date information. The fallback is deliberately conservative: it pushes ambiguous facilities into the always-treated pool rather than the clean-treated pool, attenuating the estimated treatment effect rather than inflating it.

3.1.1 Construction Timing

Figure 1 shows the distribution of data center operational years. It is heavily right-skewed: a handful of facilities date back to the 1960s–1990s, but the overwhelming majority became operational in 2005 or later. The spike at 2005 partly reflects the conservative imputation of "Existing" facilities with missing dates. Construction accelerates from 2018 onward, with 2022 and 2023 the peak years (180 and 244 new facilities). The acceleration coincides with the AI and cloud computing buildout. The figure separates pre-panel facilities (blue, operational before 2014) from within-panel facilities (red, 2014–2024), corresponding to the always-treated and clean-treated categories in our analysis. The large pre-panel stock means that most data center ZIPs are already treated when the panel begins, which limits the clean-treated sample available for the event study.

State-by-state timing distributions are in Appendix C. Virginia's boom is older and front-loaded; Texas's is more gradual, concentrated in 2019–2023.

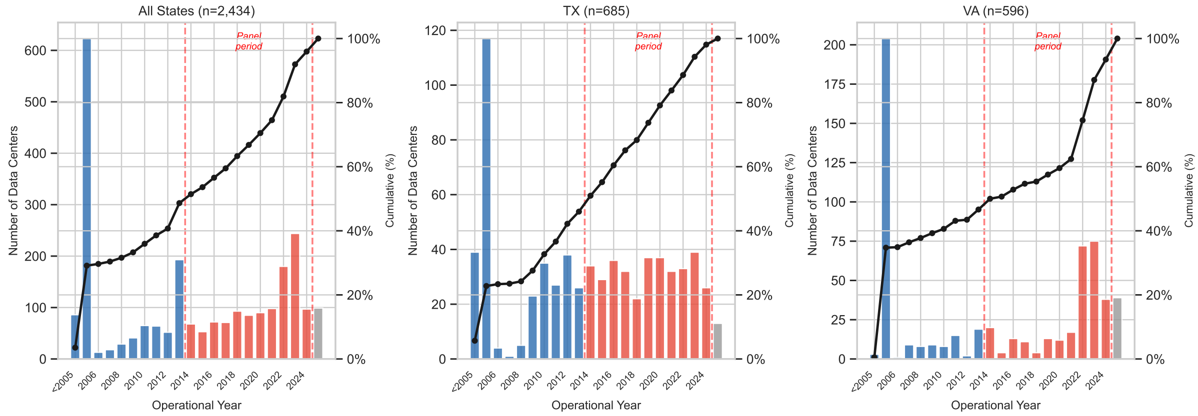

Figure 1: Distribution of data center operational years. Blue bars: pre-panel (before 2014). Red bars: within panel period (2014–2024). Black line: cumulative percentage. Red dashed lines mark the panel boundaries. Left: all states. Center: Texas. Right: Virginia. The "<2005" bin aggregates all facilities operational before 2005. Subplot labels show counts from the full scraped database ($n = 2{,}434$), which includes facilities excluded from the analytical sample of 2,277 because of post-2024 operational dates, invalid ZIPs, or missing geocoding.

3.2 Electricity Expenditures

The outcome variable is average monthly residential electricity expenditure, built from ACS PUMS for each year from 2014 through 2024, excluding 2020.[2] For each household record we use the monthly electricity cost (`ELEP`), the household sampling weight (`WGTP`), and the Public Use Microdata Area (`PUMA`) identifier; in the robustness analysis we also use household income (`HINCP`).

PUMS records are geocoded to PUMAs, which are larger than ZIP codes (about 100,000 residents each, typically encompassing multiple ZIPs). We link PUMS bills to ZIP-level data center records using the Geographic Correspondence Engine (Geocorr) from the Missouri Census Data Center (Missouri Census Data Center, 2022), with separate 2012-vintage and 2020-vintage PUMA crosswalks for the pre- and post-2022 ACS years.[3] The crosswalks give the share of each PUMA's population residing in each over-

[2] The Census Bureau did not release a standard 1-year PUMS for 2020 because of pandemic-related data collection problems.

[3] The Census Bureau redefined PUMA boundaries after the 2020 decennial census.

lapping ZIP, which we use as allocation weights:

$$\overline{\text{Bill}}_{z,t} = \frac{\sum_{i \in \mathcal{P}(z),\, t} \text{ELEP}_i \cdot \text{WGTP}_i \cdot w_{i,z}}{\sum_{i \in \mathcal{P}(z),\, t} \text{WGTP}_i \cdot w_{i,z}} \tag{1}$$

where i indexes household records, z indexes ZIP codes, t indexes years, $\mathcal{P}(z)$ is the set of PUMAs overlapping ZIP z, and $w_{i,z}$ is the PUMA-to-ZIP allocation factor.

This crosswalk-based imputation spatially smooths bills within a PUMA: ZIP codes inside the same PUMA receive estimates drawn from a common pool of records, so they are correlated and, where PUMAs contain many small ZIPs, nearly identical. The smoothing is non-classical measurement error in the dependent variable. A genuine bill increase in a treated ZIP appears in the PUMA-level mean only in proportion to that ZIP's population share, and the smoothed estimate that gets assigned back to the treated ZIP is attenuated. Symmetrically, untreated ZIPs in the same PUMA receive a spurious increase. This compresses within-PUMA treatment variation rather than just inflating standard errors. Section 7.3 discusses the implications. Aggregating to the utility-territory level in Section 5.3 mitigates the problem because territories average over many PUMAs.

3.3 Utility Service Territories

For the territory-level analysis, we map each ZIP to a utility service territory using the NREL Open Energy Information snapshot of EIA Form 861. The mapping covers 27,226 unique state-ZIP units in our 24 states across 734 utilities after filtering to the sampled states. Many ZIPs are served by more than one utility because of overlapping or dual-distribution arrangements; we keep one utility per state-ZIP, preferring investor-owned utilities (which serve most residential customers) and then using the reported residential rate only as a deterministic tie-breaker when multiple utilities remain. After merging with the bill data we have 22,086 state-ZIP units mapped to 602 utility territories. We aggregate bills to the utility-year level as a household-weighted mean across constituent state-ZIPs, using the ACS allocation weights as estimated household counts, yielding a 5,906-observation territory panel. The mapping is a 2019 snapshot; territory boundaries shift modestly over the panel period through annexations, mergers, and regulatory changes, which we treat as minor measurement error. Of the 602 territories, 92 host at least one data center at some point in the panel period; 510 never do; of the 92 treated, 55 are always-treated (DCs before 2014) and 37 are clean-treated.

3.4 Panel Construction and Descriptive Statistics

For each state-ZIP z and year t we construct: the bill $\text{bill}_{z,t}$ from Equation 1; a binary $\text{has\_dc}_{z,t}$ for any operational data center; cumulative and new counts $\text{cum\_dc}_{z,t}$ and $\text{new\_dc}_{z,t}$; and the first-DC year first_dc_year_z. State-ZIPs split into three groups by treatment timing relative to the panel: **never-treated** (n = 22,133, no DCs anywhere in the panel), **clean-treated** (n = 298, first DC in 2014–2024, the identification source for the event study), and **always-treated** (n = 403, at least one DC before 2014, used in the dose-response analysis).

Table 2: Summary Statistics

	Full Panel	Never Treated	Clean Treated	Always Treated
Avg. Monthly Bill ($)	163.83	164.03	158.80	156.56
Observations	223,409	216,447	2,956	4,006
State-ZIP Units	22,834	22,133	298	403
States	24	24	23	18
Years	10	10	10	10

Note: The panel covers 2014–2019 and 2021–2024 (2020 omitted). Bills are nominal. Observation counts reflect regression-ready rows with non-missing bills and treatment variables.

The panel is highly unbalanced: only 701 of 22,834 state-ZIP units (3.1%) have any data center presence, and data centers are concentrated in a handful of states. Bills are modestly lower in data-center state-ZIPs unconditionally, but the raw gap reflects selection (data centers locate in particular urban, suburban, and utility-service contexts) rather than a causal effect. Bills are flat from 2014 to 2019 and rise sharply from 2021 onward, consistent with post-pandemic energy price inflation and natural gas cost increases.

3.5 Data Limitations

Three data-side limitations are worth flagging upfront and matter for how the estimates should be read. First, the Interconnection.fyi inventory is incomplete: older colocation facilities permitted before interconnection tracking became common, smaller edge facilities below queue reporting thresholds, and facilities operated by vertically integrated utilities are systematically under-counted. The resulting measurement error in the treatment variable is classical and biases the estimates toward zero. Some ZIPs classified as never-treated almost certainly host a data center we have not observed, which dilutes any true effect further. Second, the PUMA-to-ZIP smoothing discussed above is non-classical measurement error in the outcome and compresses within-PUMA treatment variation, again biasing toward zero at the ZIP level. Third, the missing 2020 ACS year creates a one-year gap that complicates identification for facilities operational around 2020. Section 7.3 returns to these and other limitations.

4 Methodology

We use four causal inference strategies at the ZIP code level, supplemented by a utility-territory-level robustness analysis. Each approach addresses a different threat to identification, and the consistency (or inconsistency) of results across methods provides a form of triangulation.

4.1 Two-Way Fixed Effects Difference-in-Differences

Our baseline specification is a two-way fixed effects (TWFE) difference-in-differences model:

$$\text{Bill}_{z,t} = \alpha_z + \gamma_t + \beta_1 \cdot \text{has\_dc}_{z,t} + \varepsilon_{z,t} \tag{2}$$

where $\text{Bill}_{z,t}$ is the average monthly residential electricity bill (total household electricity expenditure) in state-ZIP z in year t, α_z is a state-ZIP fixed effect, γ_t is a year fixed effect, and $\text{has\_dc}_{z,t}$ is a binary indicator equal to one if state-ZIP z has at least one operational data center in year t. The coefficient β_1 captures the *extensive margin* effect, the average change in monthly bills associated with the presence of any data center, after absorbing time-invariant state-ZIP characteristics and common time trends. Because our dependent variable measures household electricity expenditure (price $\times$ quantity consumed) rather than the electricity rate itself (price per kWh), our estimates capture the net effect on household financial burden, which may be attenuated if households respond to rate increases by reducing consumption.

We also estimate an intensive margin specification:

$$\text{Bill}_{z,t} = \alpha_z + \gamma_t + \beta_2 \cdot \log(1 + \text{cum\_dc}_{z,t}) + \varepsilon_{z,t} \tag{3}$$

where $\log(1 + \text{cum\_dc}_{z,t})$ captures the effect of the *stock* of data centers. The log transformation accommodates the highly skewed distribution of data center counts and allows for diminishing marginal effects.

The identifying assumption underlying the TWFE estimator is the *parallel trends assumption*: in the absence of data center entry, electricity bills in treated and control ZIP codes would have evolved along parallel trajectories. The assumption is untestable, but we assess its plausibility through the event study specification described below.

Three features of our specification are worth noting. State-ZIP fixed effects (α_z) absorb all time-invariant characteristics of state-ZIPs, including geographic location, housing stock composition, baseline cost of living, state-level regulatory environment, utility service territory, and baseline rate structures. The fixed effects also handle selection on observables and time-invariant unobservables. Year fixed effects (γ_t) absorb common macroeconomic trends affecting all state-ZIPs, including nationwide energy price inflation, federal energy policy changes, and common weather shocks. We cluster standard errors at the state-ZIP level to account for serial correlation within state-ZIPs over time, following Bertrand et al. (2004).

A known limitation of the TWFE estimator in staggered adoption settings is the potential for negative weighting of certain treatment effects, as demonstrated by de Chaisemartin and D'Haultfœuille (2020) and Goodman-Bacon (2021). The problem arises when treatment effects are heterogeneous across cohorts or over time, and the TWFE estimator inadvertently uses earlier-treated units as controls for later-treated units. As Sun and Abraham (2021) showed, standard dynamic TWFE event study specifications are also subject to contamination: pre-treatment coefficients can be biased by post-treatment effects from other cohorts, and visual inspection of parallel trends can be misleading. Two features of our setting partially mitigate the problem. First, treatment (data center entry) is highly persistent: once a ZIP code receives a data center, it effectively remains treated, which reduces though does not eliminate the scope for problematic comparisons. Second, the primary event study comparison uses never-treated units as controls, avoiding the already-treated-as-control problem. We also report a cohort-size-weighted $\text{ATT}(g,t)$ difference-in-means check inspired by Callaway and Sant'Anna (2021) in Section 5.4; it is a diagnostic comparison against never-treated units, not a fully efficient Callaway–Sant'Anna estimator with formal inference.

4.2 Event Study

To assess the parallel trends assumption and characterize the dynamic pattern of treatment effects, we estimate an event study using "clean treated" state-ZIPs (those that received their first data center during the panel period) and never-treated state-ZIPs as controls:

$$\text{Bill}_{z,t} = \alpha_z + \gamma_t + \sum_{\substack{k=-5 \\ k\neq -1}}^{5} \beta_k \cdot \mathbf{1}\{t - \text{first\_dc\_year}_z = k\} + \varepsilon_{z,t} \tag{4}$$

where $\mathbf{1}\{t - \text{first\_dc\_year}_z = k\}$ is an indicator for the observation being k years relative to the first data center entry in state-ZIP z, and $k = -1$ is omitted as the reference period. For never-treated state-ZIPs, all event time indicators are zero, so these units serve as the comparison group throughout.

The coefficients $\{\beta_k\}_{k<0}$ capture pre-treatment dynamics. Under the parallel trends assumption, they should be jointly indistinguishable from zero. Statistically significant pre-treatment coefficients would suggest that treated and control ZIP codes were already on divergent trajectories before data center entry, undermining the causal interpretation. The coefficients $\{\beta_k\}_{k>0}$ capture the post-treatment dynamics and reveal whether any effect materializes immediately, builds over time, or is transitory.

Our event study includes 298 clean treated state-ZIPs and 22,133 never-treated control state-ZIPs, providing substantial statistical power for the pre-treatment diagnostic. We use a window of $k \in [-5, +5]$ to balance examining a sufficiently long pre-treatment period against retaining sample size (state-ZIPs treated early in the panel have few pre-treatment observations, while those treated late have few post-treatment observations). Observations for treated units that fall outside this window (i.e., $k < -5$ or $k > 5$) are excluded from the estimation sample to prevent them from being implicitly absorbed into the reference group, which would contaminate the baseline with distant pre-treatment and long-run post-treatment observations.

4.3 Matched Difference-in-Differences

A concern with the baseline TWFE is that treated and control state-ZIPs may differ systematically in ways that correlate with both the likelihood of receiving a data center and the trajectory of electricity bills. State-ZIP fixed effects control for time-invariant differences but do not address the possibility that treated state-ZIPs had different pre-treatment *trends*. To address this, we implement a matched difference-in-differences design that pairs each treated state-ZIP with observably similar control state-ZIPs.

The matching procedure works as follows:

1. We compute the average electricity bill for each ZIP code during the first three years of the panel (2014–2016) as a pre-treatment covariate.
2. We standardize this covariate to have zero mean and unit variance.[4]

[4] Because we match on a single scalar covariate, the standardization does not affect the relative ranking of nearest-neighbor distances. It is equivalent to matching on the unstandardized variable. We retain the standardization step for consistency with multivariate matching conventions.

3. For each clean treated state-ZIP, we compute the Euclidean distance to every never-treated state-ZIP in the standardized covariate space.
4. We select the 5 nearest neighbors for each treated state-ZIP (with replacement). Control state-ZIPs selected as matches for multiple treated units are included in the regression with frequency weights proportional to the number of times they were selected, preserving the covariate balance achieved by the matching procedure.
5. We re-estimate the TWFE DiD on the matched subsample only.

The rationale for matching on pre-treatment bill levels is that ZIP codes with similar baseline electricity expenditures are more likely to be on parallel trajectories, since they share similar housing characteristics, climate, and utility rate structures. By restricting the comparison to observably similar ZIP codes, we reduce the effective heterogeneity in the control group and make the parallel trends assumption more plausible.

One caveat is that for the subset of clean treated state-ZIPs whose first data center arrived in 2014, 2015, or 2016, the 2014–2016 average bill may partially incorporate post-treatment outcomes. This concern is mitigated by the fact that the vast majority of clean treated state-ZIPs received their first data center in 2017 or later, and the estimated treatment effects are small relative to the matching covariate variation. Future work could strengthen the design by using strictly pre-panel covariates for matching.

We report covariate balance statistics comparing treated ZIP codes, their matched controls, and the full control group, allowing assessment of whether the matching procedure achieves its objective.

4.4 Dose-Response Analysis

Our final specification uses within-state-ZIP variation in the rate of data center accumulation among "always-treated" state-ZIPs, those with at least one data center before 2014. This analysis asks a different question than the TWFE or event study: conditional on already having data centers, does the *pace* of additional construction predict bill changes?

$$\text{Bill}_{z,t} = \alpha_z + \gamma_t + \delta_1 \cdot \text{new\_dc}_{z,t} + \delta_2 \cdot \log(1 + \text{cum\_dc}_{z,t}) + \varepsilon_{z,t} \tag{5}$$

where the sample is restricted to always-treated state-ZIPs ($n = 403$), $\text{new\_dc}_{z,t}$ is the count of data centers becoming operational in year t, and $\log(1 + \text{cum\_dc}_{z,t})$ captures the cumulative stock. Both state-ZIP and year fixed effects are included.

The dose-response analysis has the advantage of comparing "like with like." All ZIP codes in the sample are data center hosts, so selection concerns are reduced. It has less statistical power than the full-sample TWFE, however, and variation in treatment intensity may be correlated with unobserved local economic conditions that also affect bills.

4.5 Utility Territory-Level Robustness Analysis

A more fundamental concern with the ZIP-code-level analysis is a potential violation of the Stable Unit Treatment Value Assumption (SUTVA). Under cost-of-service regulation, residential electricity rates are set uniformly across a utility's entire service territory through formal rate cases. If a data center in one ZIP code drives up the utility's costs, the resulting rate increase is spread across *all* ZIP codes in the service territory, not just the ZIP code where the data center is located. Adjacent ZIP codes within the same utility territory share the same rate changes regardless of whether they individually host data centers, making them invalid controls for one another.

To address the concern, we re-estimate our core specifications at the utility service territory level, the unit at which residential electricity rates are actually determined. Using the ZIP-to-utility mapping described in Section 3.3, we aggregate bills and data center counts to the utility territory (`eiaid`) level and estimate:

$$\text{Bill}_{u,t} = \alpha_u + \gamma_t + \beta_1 \cdot \text{has\_dc}_{u,t} + \varepsilon_{u,t} \tag{6}$$

where u indexes utility territories and $\text{Bill}_{u,t}$ is the average monthly residential electricity bill across all state-ZIPs in utility territory u in year t. We estimate both unweighted and household-weighted versions. The weighted version uses the sum of estimated household weights from the ACS allocation weights ($\sum \text{WGTP}_i \times w_{i,z}$) across state-ZIPs in each territory-year. We also estimate an event study and dose-response specification analogous to the ZIP-code-level models, with standard errors clustered at the utility territory level.

The utility-level analysis has 602 territories (compared to 22,834 state-ZIP units), which reduces statistical power but provides a more credible test by aligning the unit of analysis with the unit at which rates are set. Because rate changes affect all customers within a territory equally, comparing territories with and without data centers avoids the within-territory contamination that affects the ZIP-level analysis.

4.6 Inference

Throughout the ZIP-code-level analysis, we cluster standard errors at the state-ZIP level to account for serial correlation within state-ZIPs across time. Because electricity rates are set at the utility service territory level, all ZIPs within a given territory share the same rate shocks, creating within-territory correlation in the error term, a Moulton-type problem (Moulton, 1990). The theoretically ideal clustering level for the ZIP-level regressions would therefore be the utility territory (602 clusters in our mapped sample, well above the threshold for reliable cluster-robust inference). We present our main ZIP-level results clustered at the state-ZIP level, which permits arbitrary serial correlation within each state-ZIP but does not aggregate shocks across same-utility ZIPs. Clustering at the utility-territory level could yield larger standard errors if within-territory error correlation is substantial; our utility-territory-level robustness analysis (Section 5.3) addresses this concern directly by estimating at the territory level with clustering at that level. Clustering at the state level, while theoretically relevant for state-level regulatory shocks, would leave only 24 clusters, below the threshold for reliable inference (Cameron and Miller, 2015). We report clustered standard errors in parentheses in regression tables, 95% confidence intervals in event study tables, and p-values throughout. Statistical significance is noted at the 10%, 5%, and 1% levels.

5 Results

This section begins with an exploratory analysis of electricity bill patterns across our 24-state panel, then presents results from the four causal identification strategies. The exploratory analysis characterizes the data, establishes stylized facts, and examines whether covariates (income, homeownership, and state) relate to bill levels in ways that inform the causal analysis. Main-text figures show national patterns alongside the highest- and lowest-bill states; complete state-by-state figures are in Appendix D.

5.1 Exploratory Analysis

5.1.1 Electricity Bill Trends

Figure 2 plots the mean and median monthly electricity bill over time at the national level and for the highest-bill state (Georgia, \$194/mo average) and lowest-bill state (Colorado, \$138/mo average). Bills were stable from 2014 to 2019, with national mean bills fluctuating narrowly between \$144 and \$153. The stability reflects a period of low natural gas prices, moderate demand growth, and few major rate cases. Bills then rose sharply from 2021 onward, increasing from \$160 to \$200 by 2024, a 25% increase in four years. The post-pandemic surge coincided with natural gas price spikes, post-COVID demand recovery, and the first wave of rate cases incorporating pandemic-era infrastructure investments. The pattern is consistent across states: Georgia and Colorado, despite very different baseline bill levels and regulatory environments, both follow the same stable-then-surging trajectory. The commonality suggests that national energy market forces, rather than local factors like data center construction, are the primary driver of bill variation over time.

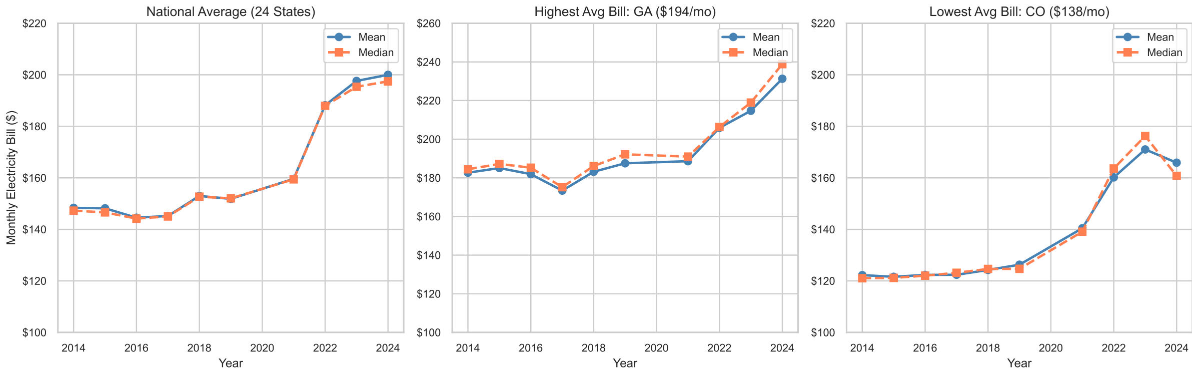

Figure 2: Mean and median monthly electricity bills over time. Left: national average across 24 states. Center: Georgia (highest average bill). Right: Colorado (lowest average bill). The gap between 2019 and 2021 reflects the missing 2020 ACS data.

5.1.2 Year-over-Year Bill Changes

Figure 3 displays year-over-year percentage changes in average bills. The national pattern shows modest declines in 2015–2016 (−0.1% to −2.6%), moderate increases in 2017–2018, and then large increases in the post-pandemic period: +5.1% in 2021, +17.9% in 2022, +5.0% in 2023, and +1.2% in 2024. The 2022 spike, the largest single-year increase in the panel, corresponds to the pass-through of sharply higher natural gas prices following the Russian invasion of Ukraine and global energy market disruptions. Georgia and Colorado both show amplified versions of the same pattern, with Georgia experiencing particularly large swings. These year-over-year dynamics indicate why year fixed effects are essential in our causal specifications: the dominant source of bill variation is macroeconomic, not local.

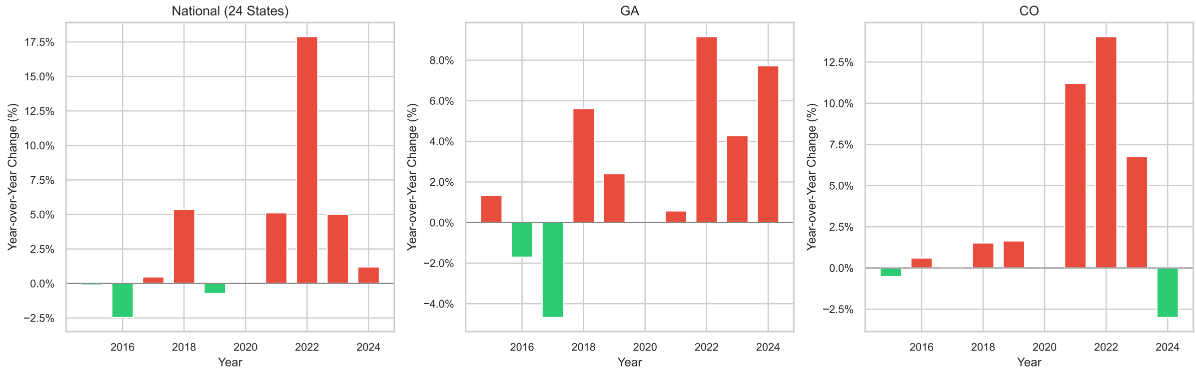

Figure 3: Year-over-year percentage change in average monthly electricity bills. Red bars indicate increases; green bars indicate decreases. Because 2020 ACS data are unavailable, the bar labeled "2021" represents the *two-year* compound change from 2019 to 2021, which may appear inflated relative to the true single-year changes. No bar is shown for 2020.

5.1.3 Data Center ZIPs vs. Non-DC ZIPs

Figure 4 compares average bill trajectories for ZIP codes with and without data centers, nationally and for the two states with the most data centers (Texas and Virginia). At the national level, DC ZIPs and non-DC ZIPs track each other closely, with DC ZIPs having slightly lower average bills in some years, the opposite of what the "data centers raise prices" narrative would predict. The pattern is even more pronounced in Texas, where DC ZIPs have consistently lower bills than non-DC ZIPs. In Virginia, the gap is also small and does not widen over time despite Virginia's massive data center expansion during this period.

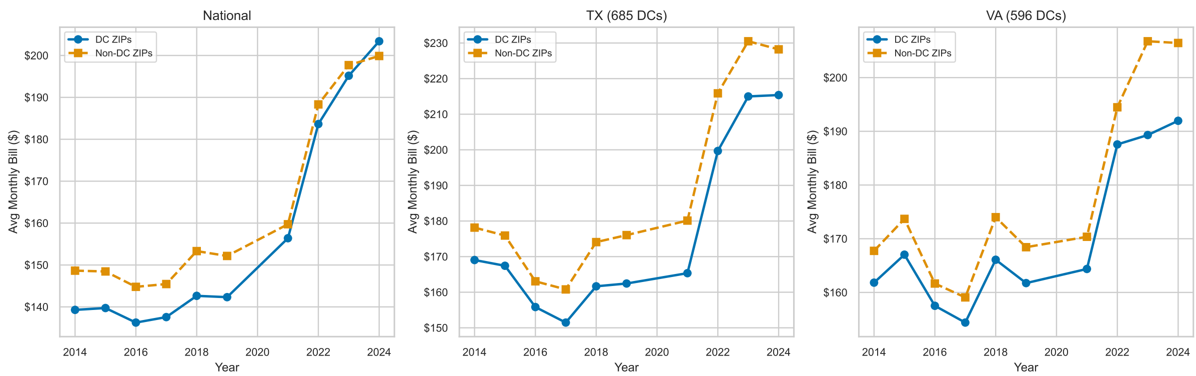

Figure 4: Average monthly electricity bills for ZIP codes with data centers (solid) versus without data centers (dashed). Left: national. Center: Texas. Right: Virginia. *Note:* Data center counts displayed in subplot labels reflect the full scraped database and may differ from the analytical sample of 2,277 reported in Table 1, which excludes post-2024, invalid-ZIP, and ungeocoded facilities.

These raw comparisons do not control for confounders and should not be interpreted causally. The absence of any visible divergence at the unconditional level, however, suggests that any causal effect is likely small.

5.1.4 Cross-State Bill Distributions

Figure 5 reports the distribution of monthly electricity bills across ZIP codes for each state in 2024. The box plots show substantial within-state heterogeneity: in most states the interquartile range spans \$30 to \$50, with some ZIP codes having average bills 2 to 3 times higher than others in the same state. The within-state variation, driven by housing stock, climate microclimates, household composition, and income, is the variation our causal analysis must parse. States also vary markedly in their median bill levels, from approximately \$130 in Colorado and Oregon to over \$220 in Georgia, reflecting differences in electricity rates, climate-driven demand, and housing characteristics.

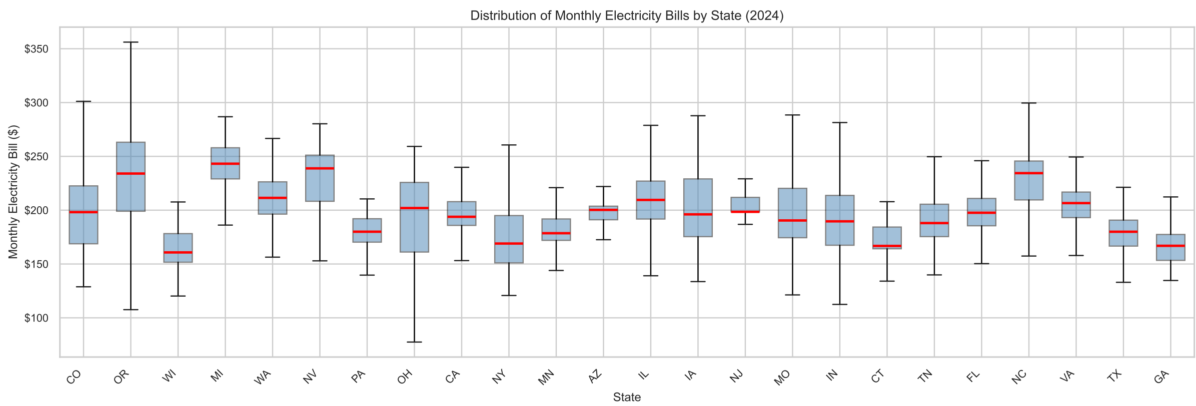

Figure 5: Distribution of average monthly electricity bills across ZIP codes by state, 2024. States ordered by median bill. Box shows interquartile range; whiskers extend to 1.5×IQR.

5.1.5 Income and Electricity Bills

To examine the relationship between household income and electricity expenditure, we aggregate raw ACS PUMS microdata for six representative states (VA, NJ, CT, OR, NV, CO) to the ZIP-year level. Figure 6 shows the cross-sectional correlation between average household income and average monthly electricity bill in 2024 for all sampled states, Virginia (highest data center count), and Connecticut (zero data centers).

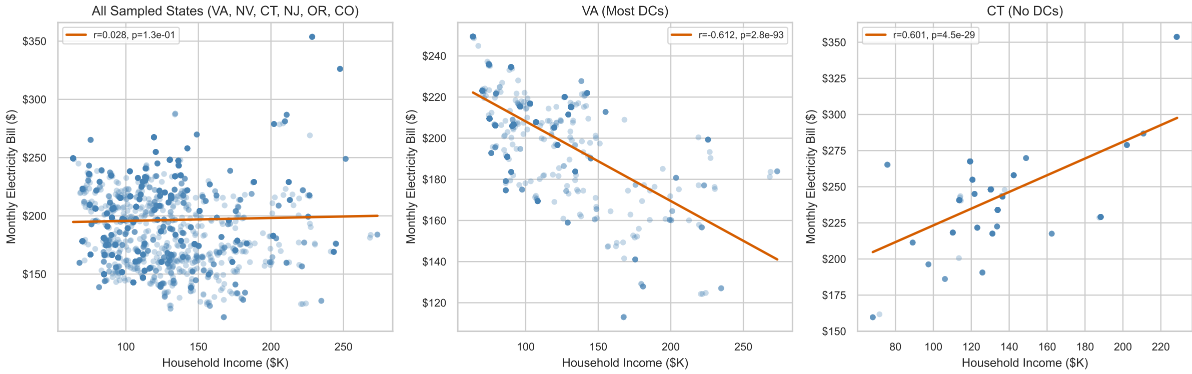

Figure 6: Scatter plot of average household income versus average monthly electricity bill at the ZIP code level (2024). Each point is a ZIP code. Red line is OLS fit with correlation coefficient r and p-value shown.

The relationship between income and electricity bills varies substantially across states and is weak in the aggregate. The overall correlation across all sampled states is near zero ($r = 0.028$), indicating that income is not a strong predictor of electricity expenditures at the ZIP code level when pooling across diverse states. The relationship is highly heterogeneous, however: some states show strong positive correlations (e.g., Connecticut), while others show negative correlations (e.g., Virginia, where $r = -0.61$), likely reflecting differences in housing stock composition, climate, and rate structures across income levels. The heterogeneity is relevant to our causal analysis: because data centers tend to locate in higher-income suburban areas, and because the income-bill relationship varies by state, absorbing location-specific effects through ZIP code fixed effects is important for isolating any treatment effect.

Figure 7 further examines whether income trends differ between DC and non-DC ZIP codes. Across the sampled states, DC ZIPs have significantly higher average household incomes than non-DC ZIPs, consistent with data centers locating in affluent suburban areas. Both groups show rising incomes over time, with no obvious divergence.

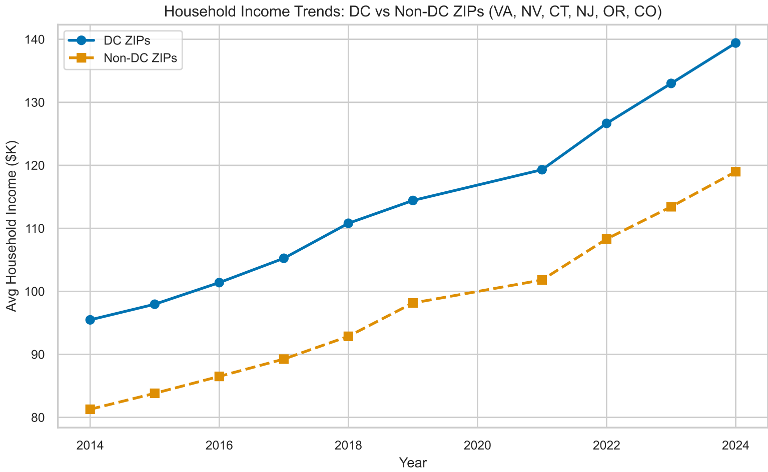

Figure 7: Average household income trends for DC ZIP codes versus non-DC ZIP codes (six sampled states: VA, NJ, CT, OR, NV, CO).

5.1.6 Homeownership and Electricity Bills

Figure 8 explores the relationship between homeownership rates and electricity bills at the ZIP code level. Homeownership is relevant because homeowners occupy larger dwellings on average, are more likely to have central heating and cooling systems, and face different rate structures than renters (who may have electricity included in rent or may occupy smaller units).

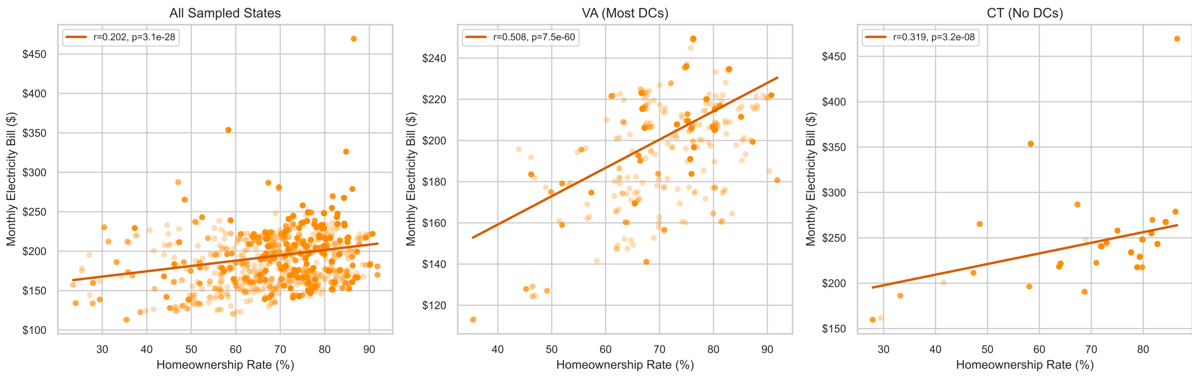

Figure 8: Scatter plot of homeownership rate versus average monthly electricity bill at the ZIP code level (2024). Each point is a ZIP code.

The correlation between homeownership rates and bills is positive: ZIP codes with higher homeownership tend to have higher electricity expenditures, reflecting larger single-family homes with greater energy demand. The relationship is consistent across Virginia and Connecticut, suggesting that it is driven by housing type rather than data center presence.

5.1.7 State-Level Correlations

Figure 9 reports a correlation matrix of state-level variables: average bill, number of data centers, data centers per ZIP code, bill change from 2014 to 2024 (in dollars and percent), and number of ZIP codes. The correlations between data center density (DCs per ZIP) and electricity bill variables are uniformly *weak*: the correlation with average bill levels is slightly negative (-0.02), while correlations with bill changes are slightly positive (0.08 for dollar changes, 0.10 for percentage changes). None of these correlations are economically meaningful. States with more data centers per ZIP code do not have systematically higher or faster-growing bills. The state-level null relationship foreshadows the ZIP-code-level null result from the causal analysis.

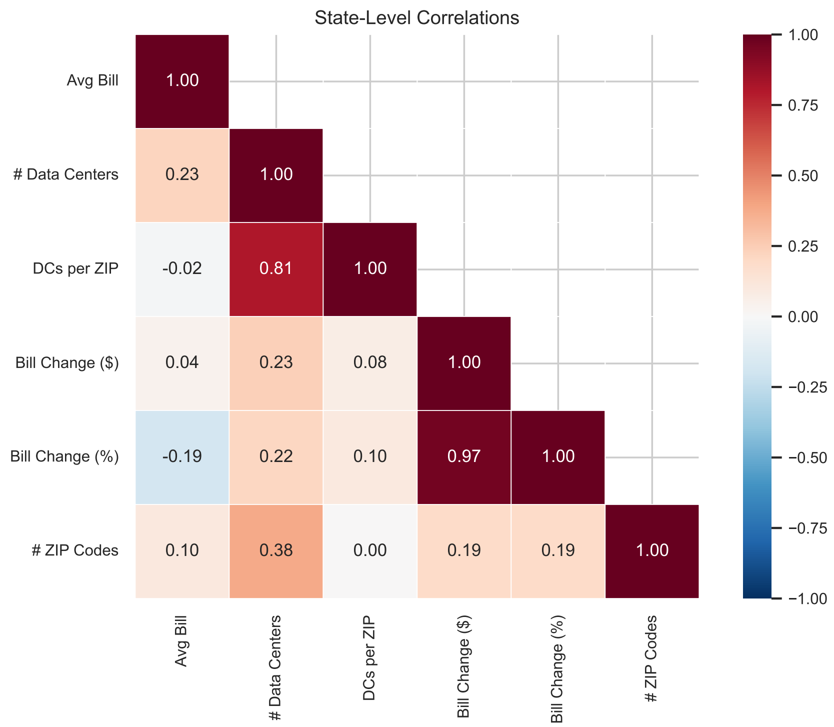

Figure 9: Correlation matrix of state-level variables. Lower triangle shows Pearson correlation coefficients.

5.2 Causal Analysis

We now turn to the formal causal analysis, beginning with the TWFE DiD.

5.2.1 Two-Way Fixed Effects DiD

Table 3 reports estimates from the TWFE DiD specifications.

Table 3: Two-Way Fixed Effects Difference-in-Differences Estimates

	(1) Extensive Margin	(2) Intensive Margin
has_dc	2.093*	
	(1.207)	
$\log(1 + \text{cum\_dc})$		0.132
		(1.320)
ZIP FE	Yes	Yes
Year FE	Yes	Yes
Clustered SE (ZIP)	Yes	Yes
Observations	223,409	223,409
R^2 (within)	0.0006	0.0000

Note: Dependent variable is average monthly residential electricity bill ($). Standard errors clustered at the ZIP code level in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

The extensive margin estimate (Column 1) associates data center presence with a $2.09 per month increase in residential electricity bills. The effect is marginally significant at the 10% level ($p = 0.083$) but does not reach conventional significance at the 5% level. The economic magnitude is modest: a $2.09 monthly increase represents approximately 1.3% of the sample mean bill of $163.83.

The intensive margin estimate (Column 2) is economically and statistically insignificant: a one-unit increase in $\ln(1 + \text{cum\_dc})$, which corresponds to multiplying the argument $(1 + \text{cum\_dc})$ by $e \approx 2.72$, or roughly a 172% increase in the data center count, is associated with a $0.13 monthly increase ($p = 0.921$). Conditional on having data centers, additional facilities do not measurably affect residential bills.

The within-R^2 values are very low (0.0001 and 0.0000), indicating that data center presence explains virtually none of the within-ZIP-code variation in electricity bills after absorbing fixed effects. This is consistent with electricity bills being driven overwhelmingly by time trends, weather, and household characteristics rather than local data center activity.

5.2.2 Event Study

Table 4 reports the event study coefficients for the 298 clean treated state-ZIPs relative to 22,133 never-treated controls.

Table 4: Event Study Coefficients

Event Time (k)	Coefficient	95% CI	Period
-5	0.62	$[-1.86, 3.10]$	Pre
-4	-0.37	$[-2.58, 1.85]$	Pre
-3	-0.06	$[-1.93, 1.81]$	Pre
-2	0.41	$[-1.20, 2.03]$	Pre
-1	0 (ref)	—	Pre
0	0.67	$[-0.98, 2.33]$	Post
1	1.83	$[-0.44, 4.11]$	Post
2	1.44	$[-1.20, 4.08]$	Post
3	2.23	$[-0.63, 5.08]$	Post
4	5.57**	$[1.61, 9.53]$	Post
5	5.22**	$[0.61, 9.82]$	Post

Note: Dependent variable is average monthly residential electricity bill ($). The sample includes 298 clean treated state-ZIPs and 22,133 never-treated controls. $k = -1$ is the omitted reference period. Standard errors clustered at the state-ZIP level. ** $p < 0.05$.

The corrected event study yields two findings. First, the pre-treatment coefficients are all small and statistically insignificant, which is more supportive of the parallel trends assumption than the earlier unwindowed specification. Second, the post-treatment profile is delayed: coefficients are positive but insignificant in the first three years after entry, then become statistically significant at $k = 4$ and $k = 5$.

The delayed-effect pattern is consistent with the hypothesis that any data center impact operates through utility rate cases, which typically introduce a lag between demand growth, capital investment, and residential rate adjustment. The interpretation remains cautious because the largest event-time effects occur at long horizons where fewer treated cohorts contribute identifying variation, and because the utility-territory-level specifications below do not show a comparable monotonic build-up.

5.2.3 Matched Difference-in-Differences

Table 5 reports covariate balance between treated state-ZIPs, their matched controls, and the full control group, and Table 6 reports the matched DiD estimates.

Table 5: Covariate Balance: Pre-Treatment Mean Electricity Bill

	Treated	Matched Controls	All Controls
Pre-treatment avg. bill ($)	141.58	141.12	147.30

Note: Pre-treatment period is 2014–2016. Matching uses 5 nearest neighbors based on Euclidean distance on the standardized pre-treatment average bill.

The matching procedure achieves good balance: the pre-treatment mean bill for treated state-ZIPs ($141.58) is nearly identical to their matched controls ($141.12), and both are lower than the overall control mean ($147.30). The pattern confirms that data center state-ZIPs, while more likely to be in urban areas, tend to have slightly lower pre-treatment bills on average, consistent with data centers locating in areas with lower electricity costs.

Table 6: Matched Difference-in-Differences Estimates

	(1) Extensive	(2) Full
has_dc	0.901	7.658**
	(1.231)	(3.205)
$\log(1 + \text{cum\_dc})$		-8.501**
		(3.782)
ZIP FE	Yes	Yes
Year FE	Yes	Yes
Clustered SE (ZIP)	Yes	Yes
Observations	16,455	16,455
R^2 (within)	0.0002	0.0016

Note: Sample restricted to matched pairs (treated state-ZIPs and their 5 nearest-neighbor controls). Controls selected multiple times enter with frequency weights. Standard errors clustered at the state-ZIP level in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

The matched extensive-margin estimate (Column 1) is \$0.90 per month, smaller than the full-sample TWFE estimate and statistically insignificant ($p = 0.464$). The attenuation after matching suggests that part of the TWFE estimate reflects residual selection: even after state-ZIP fixed effects, treated state-ZIPs differ from the average control in ways that correlate with bill trajectories.

Column 2 includes both the extensive and intensive margin variables simultaneously. The extensive margin coefficient increases substantially to \$7.66 ($p = 0.017$), while the intensive margin coefficient is negative and significant ($-\$8.50$, $p = 0.025$). The pattern of a significant positive extensive margin together with a significant negative intensive margin is difficult to interpret causally and likely reflects collinearity between the two treatment variables in the matched subsample. We view the extensive-only specification (Column 1) as more reliable.

5.2.4 Dose-Response Analysis

Table 7 reports estimates from the dose-response specification, which restricts the sample to the 403 always-treated state-ZIPs.

Table 7: Dose-Response Estimates (Always-Treated ZIP Codes)

	Coefficient
new_dc	-0.480
	(0.384)
$\log(1 + \text{cum\_dc})$	-22.300^{***}
	(3.878)
ZIP FE	Yes
Year FE	Yes
Clustered SE (ZIP)	Yes
Observations	4,006
R^2 (within)	0.0372

Note: Sample restricted to state-ZIPs with at least one data center operational before 2014. Standard errors clustered at the state-ZIP level in parentheses. *** $p < 0.01$.

The log cumulative data center coefficient is large, negative, and highly significant ($-\$22.30$, $p < 0.001$), and the new data center flow coefficient is $-\$0.48$ ($p = 0.212$). Because building a new data center in year t mechanically increases the cumulative stock `cum_dc` in the same year, the total concurrent effect of a new data center on bills combines both the flow coefficient (δ_1) and the marginal change in the log stock term ($\delta_2/(1 +$

cum_dc$_{z,t}$)). Given the large negative δ_2, the total concurrent effect of a new data center is dominated by the stock term and is therefore substantially negative. As we discuss below, however, we interpret the negative stock coefficient as reflecting mean reversion rather than a genuine causal effect.

We interpret the result with considerable caution. The negative coefficient most likely reflects *mean reversion and compositional dynamics* rather than a genuine protective effect of data center density. State-ZIPs that began the panel with very high data center counts (e.g., those in Virginia's Data Center Alley) may have experienced slower bill growth over the panel period as they approached a saturated equilibrium, while state-ZIPs with initially few data centers that then experienced rapid growth may have had other characteristics (e.g., proximity to new development) associated with faster bill increases. The within-R^2 of 0.037, while higher than in the full-sample specifications, remains low, indicating substantial unexplained variation.

5.2.5 Summary of Causal Results

Table 8 summarizes the key estimates across all four identification strategies.

Table 8: Summary of Causal Estimates

Method	Coefficient	Std. Error	p-value	N
TWFE: Extensive margin	2.093*	1.207	0.083	223,409
TWFE: Intensive margin	0.132	1.320	0.921	223,409
Event study ($k = 5$)	5.217**	2.349	0.026	—
Matched DiD: Extensive	0.901	1.231	0.464	16,455
Dose-response: new_dc	−0.480	0.384	0.212	4,006

Note: All specifications include state-ZIP and year fixed effects with standard errors clustered at the state-ZIP level. * $p < 0.10$, ** $p < 0.05$.

Across specifications, the weight of evidence does not support a strong causal effect of data centers on residential electricity bills at the ZIP code level, but the corrected event study does show delayed positive associations four to five years after entry. The baseline

$2.09 TWFE estimate is only marginally significant and is attenuated in the matched sample. The dose-response analysis provides no evidence of a positive intensive margin effect. The next subsection places these ZIP-level results on firmer methodological footing by moving to the utility territory, the level at which rates are actually set.

5.3 Utility Territory-Level Robustness Analysis

As noted in Section 4.5, a fundamental concern with the ZIP-code-level analysis is that electricity rates are set at the utility service territory level, not the ZIP code level. A data center in one ZIP code raises costs for all ZIP codes in the same utility territory, violating the SUTVA assumption that underlies the ZIP-level DiD. We re-estimate our core specifications at the utility territory level, aggregating household-weighted bills and data center counts across the 602 utility territories in the sample. Of these, 92 territories contain at least one data center, while 510 serve as never-treated controls.

5.3.1 Descriptive Comparison

Average monthly bills are similar across the two groups once state-ZIP bills are aggregated with ACS household weights. The near-equivalence is itself informative: at the level where rates are actually set, there is no large unconditional gap between DC and non-DC territories.

5.3.2 TWFE DiD at the Utility Territory Level

Table 9 reports the utility-territory-level TWFE estimates.

Table 9: Two-Way Fixed Effects DiD: Utility Territory Level

Model	Variable	Coefficient	Std. Error	p-value	N
Extensive	has_dc	0.352	(1.681)	0.834	5,906
Intensive	$\log(1 + \text{cum\_dc})$	−0.420	(1.560)	0.788	5,906
Extensive (weighted)	has_dc	−1.997	(3.863)	0.605	5,906
Intensive (weighted)	$\log(1 + \text{cum\_dc})$	−4.645	(3.095)	0.134	5,906

Note: Dependent variable is average monthly residential electricity bill ($) at the utility territory level. Utility territory and year fixed effects included. Weighted models use ACS household weights aggregated to the utility-year level. Standard errors clustered at the utility territory level in parentheses.

None of the utility-level TWFE coefficients are statistically significant at any conventional level. The unweighted extensive margin point estimate is small and positive ($0.35, $p = 0.834$), while the unweighted intensive margin is slightly negative (−$0.42, $p = 0.788$). Household weighting does not change the conclusion: the weighted extensive margin is −$2.00 ($p = 0.605$) and the weighted intensive margin is −$4.64 ($p = 0.134$). The results provide no robust evidence that data center presence raises residential bills when the analysis is conducted at the level where rates are actually determined.

5.3.3 Event Study at the Utility Territory Level

Table 10 reports the utility-level event study, estimated on the 37 clean treated territories and 510 never-treated controls.

Table 10: Event Study Coefficients: Utility Territory Level

Event Time (k)	Coefficient	95% CI	Period
-5	2.26	$[-1.72, 6.24]$	Pre
-4	0.59	$[-2.97, 4.14]$	Pre
-3	2.15	$[-1.69, 6.00]$	Pre
-2	2.34	$[-0.60, 5.27]$	Pre
-1	0 (ref)	—	Pre
0	3.57**	$[0.41, 6.72]$	Post
1	2.03	$[-1.60, 5.66]$	Post
2	1.57	$[-2.69, 5.82]$	Post
3	2.67	$[-2.68, 8.01]$	Post
4	4.51	$[-1.88, 10.89]$	Post
5	1.43	$[-4.90, 7.76]$	Post

Note: Sample includes 37 clean treated utility territories and 510 never-treated controls. $k = -1$ is the omitted reference period. Standard errors clustered at the utility territory level. ** $p < 0.05$.

The utility-level event study shows no statistically significant pre-treatment coefficients, although the pre-treatment point estimates are positive and imprecise. The absence of statistically detectable pre-treatment divergence is supportive of the parallel trends assumption, but the wide intervals reflect the smaller number of clean treated territories.

Post-treatment coefficients are positive but imprecise. The $k = 0$ coefficient is statistically significant ($3.57, 95% CI $[0.41, 6.72]$), while later horizons are not individually significant and do not show a monotonic build-up. This pattern is weaker than what would be expected if data center costs steadily accumulated into residential rates through repeated rate cases.

5.3.4 Dose-Response at the Utility Territory Level

Among the 55 always-treated utility territories, the dose-response analysis yields a significant negative coefficient on both new data centers ($-\$0.31$, $p < 0.001$) and log cumulative data centers ($-\$10.11$, $p = 0.007$). As with the ZIP-level dose-response, we attribute the negative coefficients to mean reversion (territories with the largest initial data center stocks, such as those containing Virginia's Data Center Alley, experienced slower bill growth over the panel period) rather than to a genuine protective effect of data center density.

5.3.5 Implications of the Utility-Level Analysis

The utility-territory-level results matter for two reasons. They address the SUTVA concern: by analyzing at the level where rates are actually set, we eliminate the within-territory contamination that undermines the ZIP-level approach. And the event study at the utility level shows *no* parallel trends violations, whereas the ZIP-level event study showed significant pre-treatment divergence. The cleaner identification strengthens the credibility of the null finding.

The consistency of weak-to-null TWFE results across both the ZIP-code level (22,834 state-ZIP units) and the utility-territory level (602 units) reinforces our main conclusion, while the event studies leave room for modest delayed or short-run positive effects. The utility-territory-level analysis is the more institutionally appropriate specification, since it aligns the unit of analysis with the level at which residential electricity rates are actually determined. The ZIP-code-level analysis, while offering greater statistical power, suffers from SUTVA violations (within-territory contamination) and PUMA-to-ZIP spatial smoothing that the utility-level analysis only partially mitigates. Accordingly, the utility-level TWFE null findings should be considered the primary evidence, with the ZIP-level results providing supporting context and motivation for better rate-level data.

5.4 Additional Robustness Checks

We supplement the baseline analysis with seven additional checks: (i) treatment measured by installed megawatt capacity rather than facility count; (ii) inclusion of time-varying state-ZIP-level income; (iii) subsamples by electricity market structure; (iv) hyperscaler vs. non-hyperscaler facilities; (v) a placebo test that randomly assigns treatment to never-treated state-ZIPs; (vi) a spillover test for non-DC state-ZIPs in the same utility territory as clean-treated units; and (vii) a cohort-size-weighted group-time ATT diagnostic using never-treated units as the comparison group. Results are summarized in Table 11.

Table 11: Additional Robustness Analyses

	Specification	Coef.	Std. Err.	
Baseline	TWFE has_dc	2.093*	1.207	
(i) MW exposure	TWFE cum. MW (linear, per MW)	−0.011***	0.003	
	TWFE log(1+cum. MW)	−0.611	0.392	
(ii) Income control	TWFE has_dc + log income	1.835	1.188	
	TWFE log(1+MW) + log income	−0.698*	0.388	
(iii) Market structure	TWFE has_dc | Regulated states	−0.931	1.195	
	TWFE has_dc | Restructured states	0.946	1.567	
	TWFE has_dc | Hybrid (MI only)	−1.539	1.533	
(iv) Hyperscaler	TWFE has_dc + has_hyperscaler	2.424* / −1.661	1.295 / 2.735	0.0
	TWFE cum. MW + cum. MW (hyper)	−0.028 / 0.019	0.021 / 0.023	0.1
(v) Placebo	200 random treatment assignments	0.019 (mean)	0.991 (s.d.)	
(vi) Spillover	TWFE has_dc + same-utility indicator	1.935 / −1.371***	1.207 / 0.177	0.109
(vii) Cohort DiD ATT	Average post-treatment ATT($e \geq 0$)	2.436	—	

Note: All regressions include state-ZIP and year fixed effects with standard errors clustered at the state-ZII $p < 0.05$, *** $p < 0.01$. In rows (iv) and (vi), coefficient/SE/p-values are reported in the order shown in the The (v) placebo summary reports the mean and standard deviation of the placebo coefficient across 2(assignments; 2% of placebo draws exceed the actual baseline estimate of $2.09. The (vii) cohort DiD A weighted difference-in-means diagnostic for post-treatment event times, using never-treated state-ZIPs as should not be read as a fully formal Callaway–Sant'Anna estimator with inference.

We discuss each row in turn.

(i) MW exposure. Replacing the binary presence indicator with the cumulative installed megawatt capacity in a state-ZIP (using bin midpoints, with 5 MW imputed for the 61% of facilities with unreported capacity) yields a precisely estimated *negative* coefficient: each additional MW is associated with a $0.011 decrease in the monthly bill

($p = 0.001$). The log specification is also negative but statistically insignificant. The MW-based estimate is qualitatively at odds with the binary specification, which we take as a sign that the apparent positive presence effect is not driven by the size of the installed load. Mean reversion among the small number of large-MW ZIPs (which include the long-established facilities in Loudoun County) likely contributes to the negative sign.

(ii) Time-varying income control. ACS PUMS records household income (`HINCP`); we aggregate it to the state-ZIP-year level using the same crosswalk weighting as the bill variable and add log income to the TWFE specification. For four state files with malformed ZIP-name quoting, the loader falls back to a position-based parser for the numeric fields and skips fewer than 1,100 malformed raw rows; income coverage is 99.8% of the panel. Income is strongly correlated with bills: a 1% increase in state-ZIP-year mean income is associated with a \$0.21 increase in the monthly bill ($p < 0.001$), and the within-R^2 rises from near zero to 0.162. Once income is included, the data-center presence coefficient falls from \$2.09 to \$1.84 and loses statistical significance ($p = 0.122$). The MW-based specification with income is negative and marginally significant (−\$0.70, $p = 0.072$). The reduction in the presence coefficient after controlling for income suggests that part of the baseline effect reflected time-varying income trends in data-center state-ZIPs that the fixed effects could not absorb.

(iii) Market structure. Splitting the sample by electricity market structure (regulated, restructured, Michigan's hybrid) yields point estimates of −\$0.93, \$0.95, and −\$1.54 respectively, none of which are statistically significant. The absence of a positive effect in restructured markets, where wholesale price variation should pass through more directly, is consistent with the broader null finding and does not support the hypothesis that data-center demand has been priced through to residential rates in the more competitive markets.

(iv) Hyperscaler heterogeneity. Among facilities flagged as hyperscalers (Yes or Likely) in the Interconnection.fyi metadata, neither presence nor MW-based exposure differentially raises bills. When both indicators enter together, the hyperscaler-presence coefficient is negative and insignificant ($-\$1.66$, $p = 0.54$), and the hyperscaler-MW coefficient is small and insignificant. The result is preliminary, since the hyperscaler designation is only available for the within-panel period and may be measured with error, but it does not support the view that the largest facilities have already produced detectable residential rate effects.

(v) Placebo. We randomly assign treatment status and a treatment year (drawn from the empirical distribution of clean-treated first-DC years) to 298 never-treated state-ZIPs, then re-estimate the TWFE. Across 200 random draws, the mean placebo coefficient is \$0.02 with a standard deviation of \$0.99 and a 5th–95th-percentile range of $[-1.50, 1.77]$. Only 2% of placebo draws produce a coefficient at or above the actual estimate of \$2.09, indicating that the baseline is unlikely to arise from chance assignment, but the placebo distribution is centred on zero, which weakens any inference that the \$2.09 figure reflects a structural causal effect rather than residual selection on time-varying characteristics.

(vi) Utility-territory spillover. Within-territory contamination implies that a data center's rate impact should appear not only in the host ZIP code but also in nearby ZIP codes served by the same utility. We test this directly with the ZIP-to-utility crosswalk by flagging never-treated state-ZIPs in utilities that receive a clean-treated data center elsewhere in the territory, switching their indicator to 1 from that utility's first clean-treated data-center year. The spillover indicator is negative and statistically significant ($-\$1.37$, $p < 0.001$), while the own-presence coefficient is positive but insignificant (\$1.93, $p = 0.109$). This does *not* support the earlier same-3-digit-ZIP spillover interpretation. Instead, it suggests that the small positive ZIP-level own effect is not clean evidence of utility-wide rate spreading and reinforces the need to analyze the outcome directly at the utility-territory level.

(vii) Cohort DiD ATT. As an alternative diagnostic to TWFE, we compute simple cohort-by-year difference-in-differences estimates with never-treated state-ZIPs as the comparison group and aggregate across post-treatment event times weighted by cohort size. The post-treatment average ATT is \$2.44, slightly larger than the TWFE estimate of \$2.09 but in the same range. The event-time profile (Appendix Table 14) shows small effects in the first three years after entry (−\$0.09 to \$2.09) growing to larger but more noisy effects at horizons of four years or more (\$5.97, \$4.68, \$7.45), with a few thinly populated cohorts producing pre-trend coefficients that we interpret with caution. This diagnostic does not produce qualitatively different conclusions from the TWFE: data-center entry is associated with modestly higher residential bills several years after the fact, but the effect is heterogeneous across cohorts and not strongly identified.

Implications. Taken together, the additional checks support a cautious interpretation. The MW, income-controlled, market-structure, hyperscaler, and utility-territory estimates all point to a small effect that is either insignificant or shifts toward zero (and in some specifications becomes negative) once a more plausible specification is used. The placebo result shows that the baseline positive coefficient is larger than most random assignments, but the same-utility spillover check does not support a simple rate-spreading interpretation. The picture is therefore not a clean null in every diagnostic, but it is also not robust evidence that data centers materially raised residential bills during the panel period.

6 Nowcasting Residential Bills into 2025–2026

Our causal panel ends in 2024, the year before the first hyperscale facilities were scheduled to energize. Because the American Community Survey reports ZIP-level electricity expenditure with a long lag, the bill series we use for identification will not extend into 2025 or 2026 until those survey waves are released. This section builds a simple nowcast that carries the same ZIP-level bill forward using utility rate data that are already available through early 2026. The exercise is predictive, not causal: it does not identify the effect of any data center, and we use it only to characterize where residential bills are heading and whether the near-term increases fall more heavily on ZIPs near data centers.

6.1 A rate-anchored projection

The nowcast rests on the same decomposition noted in our methodology: a household's monthly bill is a retail rate (cents per kilowatt-hour) multiplied by the quantity it consumes. Quantity is close to fixed over a two-year horizon for a given ZIP, while the retail rate is observed at the utility level and at monthly frequency in the EIA Form EIA-861M survey (U.S. Energy Information Administration, 2026), which reports residential revenue, sales, and customer counts for utilities covering most of the country through March 2026. We project each ZIP's bill by holding its most recent Census level fixed and scaling it by the change in its utility's residential rate:

$$\widehat{\text{Bill}}_{z,t} = \text{Bill}_{z,b(z)} \left(1 + \lambda \, \frac{r_{u(z),t} - r_{u(z),b(z)}}{r_{u(z),b(z)}}\right), \tag{7}$$

where $\text{Bill}_{z,b(z)}$ is state-ZIP z's bill in its most recent Census year $b(z)$ (2024 for nearly all ZIPs), $r_{u,t}$ is the EIA-861M residential rate of the utility $u(z)$ serving z, and λ is a rate pass-through coefficient. We map ZIPs to utilities with a public ZIP-to-utility crosswalk (National Renewable Energy Laboratory, 2019) whose identifiers match the EIA-861M survey; 88% of state-ZIPs, and 90% of the population we cover, map to a utility that reports monthly, and the remainder fall back to the customer-weighted average rate of the reporting utilities in their state. We estimate λ by regressing each ZIP's year-over-year Census bill growth on its utility's rate growth over 2021–2024, pooled and weighted by ZIP population, which yields $\lambda = 0.73$. A value below one is expected, since fixed monthly charges and demand response both damp the pass-through from rates to bills. To place a bill in a specific month we scale the annual projection by the utility's average seasonal shape from EIA-861M, constructed so that the twelve monthly values average to the annual level.

The level in Equation 7 is each ZIP's own recent bill, so the wide cross-sectional dispersion in bills is reproduced exactly; the projection moves that level only by the rate change the ZIP's utility has actually experienced. The price increases entering the 2025–2026 figures are therefore observed in EIA-861M rather than forecast, except for the three months past the survey's March 2026 edge, which we carry forward and flag.

6.2 Validation on held-out 2024

We assess the projection by withholding 2024: we estimate λ on 2021–2023, project the 2024 bill from each ZIP's 2023 level and its utility's 2023–2024 rate change, and compare against the actual 2024 Census bill. Across the 22,644 state-ZIPs with complete data, the projection reaches a mean absolute error of $16.7 and a mean absolute percentage error of 8.4%, with a predicted–actual correlation of 0.84 (Figure 10). A persistence benchmark that simply carries the 2023 bill forward attains a near-identical $15.6 and 7.8%. The two are statistically indistinguishable over a single year, which is what we expect: year-over-year changes in the bill are small relative to its spread across ZIPs, so last year's bill is most of the answer. The rate term earns its place over the longer horizon of the nowcast, where rate movements are larger and a frozen-2024 bill would misstate them in both directions. Both approaches improve markedly on the $26 mean absolute error of an earlier cross-sectional model that predicted bills from local housing and demographic characteristics, which we found did not generalize to held-out ZIPs.

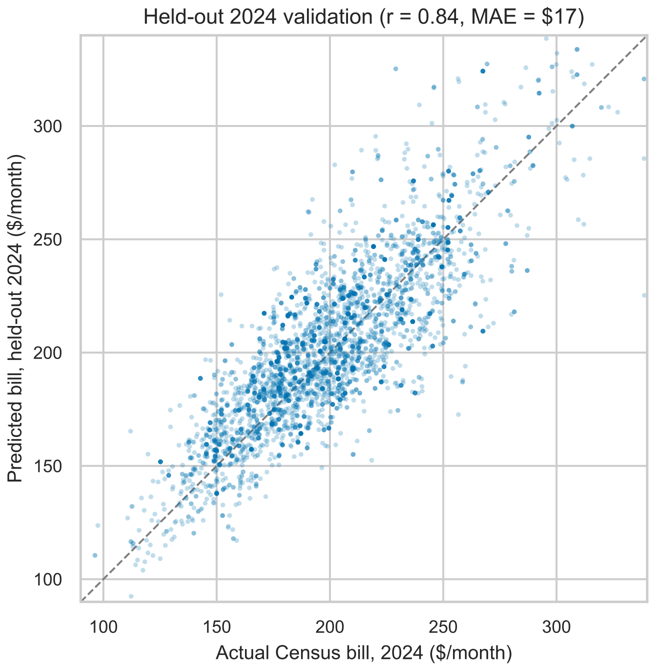

Figure 10: Held-out 2024 validation. Each point is a state-ZIP; the projection, estimated without any 2024 data, is plotted against the actual 2024 Census bill. The dashed line is the 45-degree line.

6.3 Projected bills, 2025–2026

Applying Equation 7 with λ estimated on the full 2021–2024 window, the population-weighted mean residential bill across the 24 states rises from \$199 in 2024 to \$207 in 2025 and \$209 in 2026, an increase of 5.0%. The change is uneven across states (Figure 11 and Table 12): bills rise in twenty states and edge down in four. New Jersey leads at +17%, reflecting the large 2025 Basic Generation Service auction increase, followed by Indiana, Washington, Illinois, and Pennsylvania at 9 to 13%. Nevada falls the most at −11%, with Virginia and North Carolina also slightly lower, tracking declines in those utilities' reported residential rates. These figures inherit whatever the utilities' reported rates capture, and we read them as projections that will be revised as the 2025–2026 EIA data are finalized and the corresponding ACS waves appear.

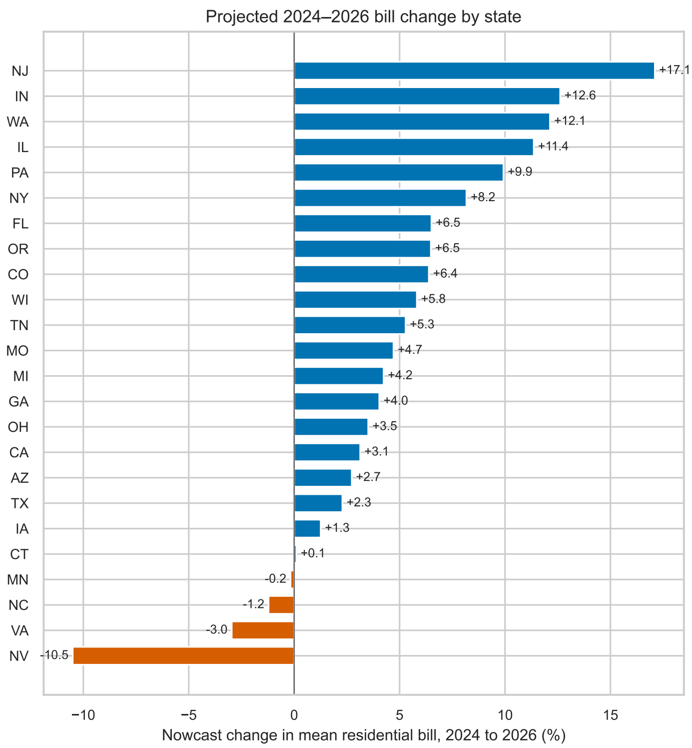

Figure 11: Projected change in the population-weighted mean residential bill from 2024 to 2026, by state.

Table 12: Projected residential bills by state, 2024–2026

State	2024 ($)	2026 nowcast ($)	Change (%)	Val. MAE ($)
NJ	199	233	+17.1	20.8
IN	188	211	+12.6	15.9
WA	172	193	+12.1	22.0
IL	163	182	+11.4	19.9
PA	193	212	+9.9	18.8
NY	202	219	+8.2	18.1
FL	213	227	+6.5	14.1
OR	173	184	+6.5	20.0
CO	156	166	+6.4	18.3
WI	172	182	+5.8	14.4
TN	204	214	+5.3	10.4
MO	183	192	+4.7	14.1
MI	179	186	+4.2	14.6
GA	216	224	+4.0	11.4
OH	180	186	+3.5	16.7
CA	221	228	+3.1	21.8
AZ	212	218	+2.7	18.9
TX	221	226	+2.3	13.3
IA	169	171	+1.3	14.7
CT	247	247	+0.1	29.3
MN	161	161	-0.2	18.0
NC	198	196	-1.2	14.3
VA	191	185	-3.0	11.3
NV	205	183	-10.5	16.5

Note: The 2024 column is the population-weighted mean Census bill; the 2026 column is the nowcast from Equation 7. "Val. MAE" is the state's mean absolute error ($/month) in the held-out 2024 validation. States are sorted by projected percentage change.

6.4 Are the increases concentrated near data centers?

Because the projection runs at the ZIP level, we can ask, descriptively, whether the near-term increases fall more heavily on ZIPs with data centers. They do not (Figure 12). The ZIP-level correlation between a ZIP's projected 2024–2026 increase and its count of operational data centers is -0.03. Grouping ZIPs by data center count, the mean projected increase is 5.8% in ZIPs with none and 1.6% in ZIPs with six or more. This is a descriptive comparison rather than a causal one, and the small high-count group is geographically concentrated, so the gradient partly reflects which utilities those ZIPs sit in. Read cautiously, it points the same way as our causal estimates: the projected near-term increases are broad and rate-driven, and are not larger where data centers cluster.

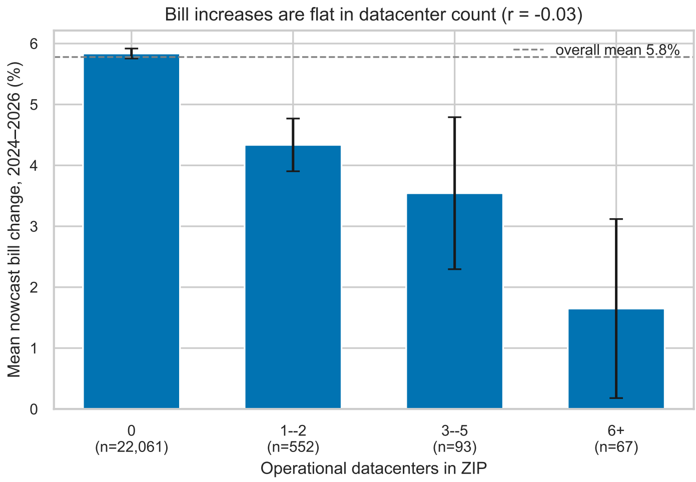

Figure 12: Projected 2024–2026 bill change grouped by the number of operational data centers in a ZIP. Bars are means with 95% confidence intervals; the dashed line is the overall mean.

6.5 Can the causal estimate be extended to 2025–2026?

A natural question is whether these projections can update the paper's causal estimate through 2025–2026. They cannot, for a reason worth stating. Using the nowcast itself as the outcome would be circular: Equation 7 contains no datacenter term, so a difference-in-differences on predicted bills would mechanically return zero plus whatever selection links datacenter location to utility rate trajectories. The alternative is to run the difference-in-differences on the *actual* utility rates, which EIA-861M reports through early 2026, at the utility-territory level where the paper found its most reliable estimates. The recent window offers almost no identifying variation, however: of the 186 reporting utility territories in our states, only two gain their first datacenter during 2021–2025, since most datacenter territories were already treated by 2020 and 76% of facilities lack a recorded operational date. A two-way fixed-effects regression of the residential rate on datacenter presence over 2020–2025 returns a non-positive coefficient (customer-weighted, about -1.6 cents per kWh), in line with the paper's pre-2024 results, but it is identified off those two transitions and is confounded by the tendency of datacenters to locate where power is cheap and rates are flat. We therefore do not report it as a causal estimate. The bind is structural: recent outcomes are observed only at the utility level, where datacenter treatment barely moves, while the granular ZIP-level treatment the paper exploits has no observed 2025–2026 outcome until the corresponding ACS waves are released. A credible causal update must wait for those data. What the projection does support is the descriptive reading above, that the near-term increases are broad and rate-driven and are not larger where datacenters cluster.

6.6 Limitations

The nowcast carries several caveats beyond its non-causal nature. The EIA-861M data for 2025 and 2026 are marked preliminary and will be revised, so the projection is refreshed on each release. The pass-through λ is calibrated on four Census years of relatively modest rate movement, and applying it to the larger 2024–2026 changes is an extrapolation that the held-out test cannot fully confirm. There is no ZIP-level monthly ground truth, so the monthly seasonal pattern is taken from the utility's EIA series rather than validated at the ZIP level. Finally, the ZIP-to-utility crosswalk is a 2019 vintage, and a minority of ZIPs rely on a state-level rate where their own utility does not report monthly. None of these affects the qualitative reading that projected near-term increases are broad and not concentrated near data centers.

7 Discussion

Our results, weak-to-null effects across multiple identification strategies at both the ZIP code and utility territory levels, may initially seem at odds with the intensity of public concern about data centers and electricity costs. In this section, we discuss potential explanations for the findings, their implications for policy, and the limitations of the analysis.

7.1 Why Might Data Centers Not Affect Local Residential Bills?

We identify five mechanisms that could explain the absence of a strong local effect.

7.1.1 Regulated Rate Structures

The most plausible explanation is the institutional structure of electricity pricing in the United States. In most of the 24 states in our sample, residential electricity rates are set through formal rate cases before state public utility commissions (PUCs). Under cost-of-service regulation, a utility's total revenue requirement, including generation, transmission, and distribution costs, is allocated across customer classes (residential, commercial, industrial) according to cost causation principles and policy judgments about affordability. Rates are uniform within a customer class across the utility's entire service territory, which typically spans multiple counties or an entire state.

This means that even if data center demand drives up the utility's total costs (e.g., by necessitating new generation capacity or transmission upgrades), the costs are spread across *all* residential customers in the service territory, not just those in ZIP codes near data centers. A data center that increases Dominion Energy Virginia's capital expenditures would raise rates for all 2.7 million Dominion residential customers, not just those in Loudoun County. At the utility-territory level, the per-customer impact is highly diluted. Our utility-territory-level robustness analysis (Section 5.3) directly tests whether territories hosting data centers experience higher bills than territories without, and finds no significant effect. The null result at the territory level indicates that the aggregate costs data centers impose on host utilities are sufficiently small, when spread across the full customer base, to be undetectable in our data. The distinction matters here: rate-spreading operates *within* a territory to distribute costs uniformly, but the territory as a whole still bears the full aggregate cost. The null territory-level finding reflects cost dilution across a large customer base, not cost elimination.

7.1.2 Separate Tariff Schedules

Data centers typically receive electricity under large commercial or industrial tariff schedules, often with demand charges, high-voltage delivery, and interruptibility provisions that differ substantially from residential rate structures. In many jurisdictions, utilities have created dedicated tariff categories for data centers and other large loads (e.g., Dominion Energy's Schedule GS-4 for demands above 500 kW). The cost allocation between these tariff classes is a regulatory decision made during rate cases, and PUCs have generally sought to prevent cross-subsidization, ensuring that large commercial customers pay their fair share of system costs.

If the regulatory process effectively isolates residential ratepayers from the direct costs of serving data centers, then data center presence would not affect residential bills even at the utility-territory level. While there is ongoing debate about whether current cost allocation methodologies fully capture the system costs imposed by data centers (particularly peak demand impacts and transmission congestion), the regulatory intent is to prevent cost-shifting.

7.1.3 Grid-Level vs. Local-Level Effects

Our analysis tests for effects at the ZIP code level, but the channels through which data centers might affect electricity prices operate at much larger geographic scales. Electricity generation costs are determined by regional wholesale markets (e.g., PJM, ERCOT, MISO) that span multiple states. Transmission costs are allocated at the utility or regional transmission organization (RTO) level. Distribution costs, while more localized, are still aggregated across utility service territories.

A ZIP-code-level analysis is well-suited to detect *local* effects, for example if data centers caused localized distribution congestion that required targeted infrastructure upgrades borne by nearby ratepayers. But the primary cost channels (new generation capacity, transmission expansion, higher wholesale energy prices) operate at regional or utility-territory scales. Our national year fixed effects (γ_t) absorb only *common national* time trends, not region-specific shocks. If data center demand raises wholesale prices in specific regional markets (e.g., PJM or ERCOT) where data centers are disproportionately concentrated, these region-specific cost increases would remain in the error term and could positively correlate with the treatment indicator. Including state-by-year or region-by-year fixed effects would address the concern but at the cost of absorbing much of the identifying variation. Our utility-territory-level analysis (Section 5.3) partially addresses this by moving the unit of analysis one level up, and still finds null effects, suggesting that even at the territory level, data center cost pass-through to residential rates is negligible during the period studied. Effects operating at the regional wholesale market level, which neither our ZIP nor our utility analysis can fully isolate, remain a possibility.

7.1.4 Endogenous Location Choice

Data centers are large infrastructure investments that optimize location based on, among other factors, the cost and availability of electricity. They tend to locate in areas with competitive electricity rates, capable grid infrastructure, and favorable regulatory environments. The endogenous selection creates a *negative* correlation between data center presence and electricity costs that could offset any causal positive effect. Virginia's data center boom, for example, was partly driven by the state's historically moderate electricity rates and Dominion Energy's willingness to accommodate large loads. If data centers select into areas with low and stable electricity prices, the observed correlation between data center presence and bill levels may be zero or negative even if data centers exert upward pressure on costs.

Our TWFE specification addresses this concern through ZIP code fixed effects, which absorb the time-invariant component of location selection. If selection operates on time-varying factors, however (for example, if data centers locate in ZIP codes experiencing relatively slower bill growth), the fixed effects strategy would not fully address endogeneity.

7.1.5 Data Measurement Limitations

Several features of our data reduce our ability to detect true effects. The PUMA-to-ZIP crosswalk introduces measurement error in our outcome variable by spatially smoothing bills across ZIP codes within the same PUMA. As discussed in Section 3, this constitutes non-classical measurement error that compresses within-PUMA variation in the dependent variable, inflating standard errors and reducing statistical power. While classical measurement error in the dependent variable does not bias point estimates, the spatial smoothing in our setting may produce attenuation if treated and untreated ZIP codes share PUMAs. Our data center inventory may also be incomplete, missing smaller facilities that predate comprehensive interconnection tracking; the resulting measurement error in the *treatment* variable would cause classical attenuation bias, pushing point estimates toward zero.

7.2 Implications for Policy

Our findings have several implications for the ongoing policy debate about data centers and electricity costs.

7.2.1 Local Price Effects May Be Small

The most direct policy implication is that the frequently cited concern, "data centers are raising my electricity bill," does not find support in our data at either the ZIP code level or the utility territory level. This does not mean the concern is unfounded: data center demand growth *could* raise electricity costs at the regional wholesale market level through channels our analysis cannot detect. But the localized version of the claim, that living near a data center, or even being served by a utility that hosts data centers, specifically raises one's bills, is not supported by our evidence.

7.2.2 Aggregate Effects Remain an Open Question

Our null findings at both the ZIP code and utility territory levels are consistent with the possibility that data center demand growth raises electricity costs at the aggregate level.

If data center load growth drives up wholesale energy prices, necessitates expensive new generation capacity, or accelerates transmission investment, the costs would be spread across all ratepayers in a regional market (e.g., PJM, ERCOT). To the extent that the cost increases are correlated across regions and manifest as a national trend, they would be partially absorbed by our year fixed effects. Region-specific wholesale price increases driven by locally concentrated data center demand, however, would not be absorbed by national year fixed effects and could appear as a positive bias in our treatment coefficient, making our null findings conservative with respect to this channel. Future research using regional wholesale price data or multi-state market-level analysis could investigate the aggregate channel.

7.2.3 Regulatory Frameworks May Be Functioning as Intended

From a regulatory perspective, our findings can be read positively: the system of cost-of-service regulation and tariff class separation appears to be effective at insulating residential ratepayers from the direct costs of serving large industrial loads. The regulatory framework, while imperfect, seems to provide a degree of rate stability for residential customers even as data center demand grows.

7.2.4 Long-Run Effects and Infrastructure Investment

Our panel covers only 10 years (2014–2024), a period that may be too short to capture the long-run effects of data center-driven infrastructure investment. Major generation and transmission projects have lead times of 5 to 15 years, and the rate impacts of capital investments are amortized over decades. The recent surge in data center construction (2022–2025) may generate rate impacts that only become visible in the 2025–2030 period as utilities complete rate cases that incorporate the associated capital expenditures. Our analysis captures the first decade of the data center boom but may miss the delayed pass-through of infrastructure costs.

7.3 Limitations

We acknowledge several limitations of the analysis.

Incomplete data center inventory. Our inventory of 2,277 analytical-sample data centers, drawn from Interconnection.fyi's aggregation of interconnection queue filings, is almost certainly a substantial undercount of the true population of data centers in the United States. Industry estimates place the total number of U.S. data centers well above this figure; our sample likely captures roughly half or fewer of all operational facilities. The interconnection queue methodology systematically misses several categories of facilities: older colocation and enterprise data centers that were built before interconnection tracking became widespread; smaller edge data centers that fall below queue reporting thresholds; facilities operated by vertically integrated companies that do not appear in public filings; and data centers in states or utility territories with less transparent interconnection processes. The undercount introduces measurement error in our treatment variable that biases our estimates toward zero through classical attenuation. More importantly, it means that some ZIP codes classified as "never-treated" in our analysis may in fact contain data centers we have not identified, which would further dilute any estimated treatment effect. To the extent that unobserved data centers are systematically different from observed ones, for example if smaller facilities are more likely to be missed and these facilities have different cost profiles or locational patterns, the bias could be non-classical. Future research should triangulate data center locations using multiple sources (interconnection queues, commercial real estate databases, satellite imagery, building permit records) to construct a more complete inventory.

Temporal coverage and the hyperscaler wave. Our panel ends in 2024, just before the largest wave of data center construction is expected to come online. The period from 2025 onward marks a qualitative shift in the scale and nature of data center development: hyperscale facilities planned by Amazon Web Services, Google, Microsoft, and Meta, many exceeding 500 MW and some approaching or surpassing 1 GW of capacity, are under construction across Virginia, Texas, Georgia, Ohio, and other states. These facilities are an order of magnitude larger than the typical data centers in our sample and will impose fundamentally different demands on the grid, potentially requiring dedicated generation assets, major transmission upgrades, and new substation construction. The absence of statistically significant price effects through 2024 should therefore *not* be read as evidence that data center growth will not affect residential electricity prices going forward. The cost impacts of the hyperscaler buildout, including the capital expenditures for new generation and transmission that will be recovered through utility rate cases over 20- to 40-year horizons, will likely begin appearing in residential rates only in the late 2020s and 2030s. Our analysis captures the era of moderate, incremental data center growth; the era of larger, concentrated hyperscale deployment remains unexamined and may produce qualitatively different outcomes for residential ratepayers.

Missing 2020 data. The absence of 2020 ACS data creates a one-year gap in our panel that may confound the identification of treatment effects for data centers that became operational around 2020. The post-2020 period also coincides with significant energy market disruptions (natural gas price spikes, post-pandemic demand recovery) that may introduce structural breaks in bill trajectories.

Imputed operational dates. For a substantial fraction of data centers, we lack precise operational dates and must impute them from approval or application dates. If the true operational date differs from the imputed date by several years, this introduces measurement error in the timing of treatment that goes beyond simple attenuation: mistimed entry creates misleading lead and lag patterns in the event study that can distort both pre-treatment trend tests and post-treatment effect estimates.

Presence-based rather than capacity-based treatment. Our treatment variables measure data center *presence* (binary) and *count* (cumulative number of facilities), but the economic channel through which data centers might affect electricity costs operates through electrical *load* (megawatts of demand). Our interconnection queue data includes MW capacity for many facilities, and substituting or supplementing the current count-based treatment with MW-based measures would provide a tighter link between the econometric specification and the underlying economic mechanism. A 500 MW hyperscale facility imposes qualitatively different grid demands than a 5 MW colocation center, and treating them identically dilutes the signal.

Limited covariates. Beyond the time-varying ZIP income added in Section 5.4, we do not control for other household-side covariates that might co-move with data center entry: housing stock composition, square footage, appliance penetration, or weather. ZIP fixed effects absorb the time-invariant component of these characteristics, and the year fixed effects absorb nationwide weather and price shocks, but residual variation in housing turnover or local climate could still confound the estimates. Census Summary File data on housing characteristics at the ZCTA level would address this if combined with PRISM weather data; we leave the extension to future work.

PUMA-to-ZIP spatial smoothing. A fundamental limitation of our ZIP-code-level analysis is that ACS PUMS microdata identifies household geography only at the PUMA level (areas of approximately 100,000 residents, typically encompassing multiple ZIP codes). Our crosswalk-based imputation of ZIP-level bills from PUMA-level data produces spatially smoothed estimates: ZIP codes within the same PUMA have correlated outcome measures, and genuine within-PUMA treatment variation is attenuated. This constitutes non-classical measurement error that biases ZIP-level treatment effects toward zero and may compromise the matched DiD strategy if matched controls are drawn from the same PUMA as treated units (yielding mechanically similar outcome trajectories). Future work using ACS Summary File data at the ZCTA level (which provides true ZIP-code-tabulation-area aggregates) or individual utility billing data would overcome the limitation. Our utility-territory-level analysis partially mitigates this concern by aggregating to a coarser unit that averages over many PUMAs.

Geographic unit. As discussed above, the ZIP code may not be the natural geographic unit at which electricity pricing responds to demand. Our utility-territory-level robustness analysis (Section 5.3) addresses this concern by re-estimating at the level where rates are set, and finds consistent null results. Effects operating at the regional wholesale market level (e.g., PJM, ERCOT, MISO), however, would not be fully absorbed by national year fixed effects at either the ZIP or utility levels.

Heterogeneous regulatory environments. Our 24 states span a range of regulatory structures, from fully regulated (e.g., Virginia, Georgia) to restructured (e.g., Texas, Ohio, Illinois). Pooling across these environments may mask heterogeneous effects that are detectable in some regulatory contexts but not others. Subgroup analyses by market structure could be informative.

Limited clean treated sample. Only 298 state-ZIPs and 37 utility territories received their first observed data center during the panel period, limiting the statistical power of the event studies. The ZIP-level event study does estimate positive effects at longer horizons ($k = 4$: $[1.61, 9.53]$; $k = 5$: $[0.61, 9.82]$), but these cells are identified by fewer cohorts and should be interpreted with caution. At the utility level, the corresponding longer-horizon confidence intervals are wide ($k = 4$: $[-1.88, 10.89]$; $k = 5$: $[-4.90, 7.76]$), so we cannot rule out economically meaningful positive effects with the available sample.

Parallel trends and event-study interpretation. After excluding treated observations outside the $[-5, +5]$ event window, the ZIP-level event study no longer shows statistically significant pre-treatment coefficients. This is an improvement over the earlier unwindowed specification, but it does not eliminate all identification concerns: long-horizon post-treatment effects rely on thinner cohorts, and staggered-adoption TWFE event studies can still be contaminated by heterogeneous treatment effects. The utility-territory-level event study also shows no significant pre-treatment coefficients, but its post-treatment estimates are imprecise and do not show a monotonic pattern. We therefore interpret the event studies as suggestive diagnostics rather than definitive dynamic causal estimates.

No causal mediator analysis. Our analysis identifies the reduced-form relationship between data center presence and bills but does not decompose the channels through which effects might operate (wholesale price impacts, capacity charges, transmission costs, distribution upgrades). A structural analysis that traces the cost pass-through chain from data center demand to residential rates would provide richer policy guidance.

7.4 Future Research Directions

The limitations above point toward several avenues for follow-up work.

Individual-level electricity bill analysis. The most useful extension would be to obtain *individual household-level* electricity bill data rather than relying on Census-aggregated ZIP code averages. Our current approach, aggregating ACS PUMS microdata to the ZIP code level via PUMA-to-ZIP crosswalks, introduces substantial measurement error and obscures individual-level variation. ZIP code averages conflate households with very different energy consumption profiles: a 4,000-square-foot home with central air conditioning will respond differently to rate changes than a 900-square-foot apartment, and these housing types may be distributed differently in data center ZIP codes versus control ZIP codes. Individual-level data could be obtained through partnerships with electric utilities, which maintain detailed billing records for every customer. A matched comparison design using individual bills would be considerably more powerful: one could match homes by square footage, building age, heating/cooling type, occupancy, and baseline electricity consumption, then compare bill trajectories for matched homes across utility territories that differ in data center exposure. Because residential electricity rates are uniform within a utility's service territory, individual-level comparisons within the *same* utility territory would capture differences in energy consumption patterns rather than rate differences. Comparisons *across* utility territories would more directly test for rate effects. Frameworks such as the California Energy Commission's Energy Data Access Committee and New York's Utility Energy Registry can support work of this kind. An additional advantage of individual-level data is the ability to examine *distributional* effects: even if average bills do not change, data center development could raise bills for low-income households (who are less likely to have energy-efficient homes and may face higher marginal rates under tiered pricing) while lowering bills for high-income households who benefit from grid infrastructure improvements.

Pre-politicization qualitative research. Once a policy issue becomes salient and partisan, survey responses and behavioral data become contaminated by motivated reasoning, media framing effects, and social desirability bias. The window for collecting unbiased baseline perceptions about data centers and electricity costs is closing as media coverage intensifies. A multi-site qualitative study could conduct structured interviews with residential ratepayers in communities at various stages of data center development (from areas where construction has just been announced, through areas with established but not yet controversial facilities, to areas where the issue has become politically contentious), document households' *unprompted* attributions for electricity bill changes, survey residents' willingness to accept data centers under various compensation scenarios before these scenarios become associated with partisan positions, and track how perceptions evolve longitudinally as data centers are announced, constructed, and become operational. The evidence would complement our quantitative analysis by illuminating how data centers affect *perceived* (as opposed to actual) electricity costs.

Utility rate case analysis. A complementary approach would be a structural examination of utility rate cases in states with significant data center growth. When utilities like Dominion Energy Virginia, Georgia Power, or Portland General Electric file rate cases with their respective PUCs, the filings contain detailed information about the cost drivers underlying proposed rate increases, including load growth projections, capital investment plans, and cost allocation methodologies. A systematic content analysis of rate case filings could quantify the share of proposed capital investments attributable to data center load growth versus other demand drivers, trace the regulatory process by which data center-driven costs are allocated across customer classes, identify whether PUCs have modified cost allocation rules in response to data center growth, and compare rate impacts in states with proactive data center tariff policies versus states without.

Grid congestion and reliability analysis. Our analysis focuses exclusively on electricity prices, but data centers may affect residential welfare through non-price channels, particularly grid reliability. Anecdotal reports from Oregon, Virginia, and Texas suggest that data center concentration can cause local grid congestion, leading to more frequent brownouts, voltage fluctuations, and delayed interconnection for new residential construction. Merging our data center location data with utility-level reliability metrics from EIA Form 861 (SAIDI, SAIFI, CAIDI) would test whether utilities hosting data centers experience worse aggregate reliability. EIA Form 861 reports reliability data at the utility and state level rather than at the ZIP code level, so testing for *localized* reliability impacts within specific ZIP codes would require granular outage management system (OMS) data obtained directly from utilities or third-party outage tracking services.

Water consumption and environmental justice. Data centers consume significant quantities of water for cooling, particularly in hot climates. Mytton (2021) estimated that a typical hyperscale data center consumes 1 to 5 million gallons of water per day. In water-scarce regions like Arizona, Nevada, and parts of Texas, the consumption may compete with residential, agricultural, and ecological water needs. An environmental justice analysis could examine whether data centers disproportionately locate in communities of color or low-income communities, whether these communities bear disproportionate environmental costs (noise, water depletion, diesel generator emissions), and whether community benefit agreements and tax incentive packages adequately compensate for these externalities.

Natural experiments from policy changes. Several recent state-level actions could serve as natural experiments. The Virginia General Assembly considered but ultimately rejected proposals to pause data center tax incentives in 2024–2025, creating variation in policy expectations that could affect both data center investment and residential perceptions. Georgia Power introduced a new tariff structure for data centers in 2024 that includes demand charges and capacity reservation fees; comparing residential rate trajectories before and after this reform could isolate the rate-shielding effect of tariff design. ERCOT's 2024 rules requiring large loads (including data centers) to participate in demand response could affect wholesale prices and, by extension, retail rates. The Oregon PUC's investigation into how data center growth costs should be allocated across customer classes provides a real-time experiment in regulatory design. Exploiting these policy changes through difference-in-differences or regression discontinuity designs would yield cleaner identification of the causal mechanisms linking data center policy to residential electricity costs.

Long-run dynamic analysis. Our 10-year panel may be too short to capture the full dynamic of data center impacts on electricity infrastructure and rates. The electricity system operates on multiple timescales: wholesale energy prices adjust hourly, fuel costs fluctuate monthly, utility rate cases occur every 1 to 5 years, and major infrastructure investments (generation plants, transmission lines) have lead times of 5 to 15 years and cost recovery periods of 20 to 40 years. A longer-run analysis, extending back to the early 2000s when the first large data center clusters emerged in Virginia and the Pacific Northwest, could capture whether the capital investments driven by data center growth have been reflected in residential rates over longer horizons. The exercise would require assembling historical rate case data and utility billing records, which is feasible but labor-intensive.

Cross-country comparative analysis. The United States is not the only country experiencing rapid data center growth. Major data center hubs exist in Ireland (where data centers consume approximately 18% of national electricity), the Netherlands, Singapore, and the Nordic countries. These jurisdictions have very different electricity market structures, regulatory frameworks, and pricing mechanisms, creating variation that could help identify the conditions under which data center growth does or does not affect residential electricity prices. Ireland is a useful case: the country's grid operator, EirGrid, has imposed a moratorium on new data center connections in the Dublin area due to capacity constraints, and the Commission for Regulation of Utilities has explicitly linked data center demand to residential rate increases. A comparative analysis of the Irish and U.S. experiences could illustrate how market structure, regulatory capacity, and grid size mediate the relationship between data center growth and residential electricity costs.

8 Conclusion

This paper investigates whether data centers cause increases in residential electricity prices, using a state-ZIP-level panel dataset of 22,834 units across 24 U.S. states for 2014–2024. We link 2,277 analytical-sample data centers, identified from interconnection queue filings, to household electricity expenditure data from the American Community Survey, and apply four causal inference strategies (two-way fixed effects difference-in-differences, event study analysis, matched difference-in-differences, and dose-response estimation) supplemented by a utility-territory-level robustness analysis that aligns the unit of analysis with the level at which electricity rates are actually determined.

Findings are mixed but do not provide robust evidence that data center presence causes meaningful increases in residential electricity bills. At the ZIP code level, the baseline TWFE extensive margin estimate is \$2.09 per month and only marginally significant ($p = 0.083$), while the matched DiD estimate attenuates to \$0.90 ($p = 0.464$). The event study shows delayed positive coefficients four and five years after data center entry, but these delayed associations are not mirrored by the more institutionally appropriate utility-level TWFE models. At the utility territory level, where rates are actually set, the unweighted TWFE estimate is \$0.35 ($p = 0.834$) and the household-weighted estimate is $-$\$2.00 ($p = 0.605$). The dose-response analysis at both levels finds no evidence that additional data center construction raises bills.

A battery of additional robustness checks reinforces the cautious interpretation. Replacing the binary presence indicator with the cumulative installed megawatt capacity flips the sign and produces a small but precisely estimated negative coefficient ($-$\$0.011/MW, $p = 0.001$). Adding a time-varying household income control drops the presence coefficient to \$1.84 ($p = 0.122$) and raises the within-R^2 from near zero to 0.162, indicating that state-ZIP-year income absorbs part of the apparent baseline effect. Subsample estimates by market structure (regulated, restructured, hybrid) are all small and insignificant. Hyperscaler facilities do not differentially raise bills relative to colocation centers. A cohort DiD ATT diagnostic returns an aggregate post-treatment effect of \$2.44, in the same range as the TWFE estimate. A placebo test with random treatment assignment produces a coefficient distribution centred on zero (s.d. \$0.99). A same-utility spillover test does not support the earlier same-3-digit-ZIP spillover interpretation: the spillover coefficient is negative ($-\$1.37$, $p < 0.001$), while the own-presence coefficient is positive but insignificant.

These weak-to-null results should be interpreted with two caveats that constrain their generalizability. First, the inventory of 2,277 analytical-sample data centers, sourced from interconnection queue filings, likely captures only a subset of all operational U.S. data centers, as many older, smaller, or privately operated facilities are not reflected in public interconnection records. The undercount biases our estimates toward zero and means some ostensibly "untreated" ZIP codes may host unobserved data centers. Second, the panel ends in 2024, just before the largest wave of data center construction comes online. The hyperscale facilities of 500 MW to 1 GW+ planned by major cloud and AI companies for 2025 onward are an order of magnitude larger than the data centers in our sample. The findings reflect the era of moderate, incremental data center growth and should not be extrapolated to the hyperscaler era, whose grid-level cost impacts will likely materialize in residential rates only over the late 2020s and 2030s as utilities complete rate cases incorporating the associated capital expenditures.

Our results do *not* imply that data centers have no effect on electricity costs. They indicate that any such effects do not manifest as differences in household electricity *expenditures* at either the ZIP code or utility territory level within the pre-hyperscaler period we study. The dependent variable is also worth noting: we measure total monthly electricity bills (price $\times$ quantity), not electricity rates (price per kWh). If data centers cause electricity rates to increase, households may partially offset higher prices by reducing consumption, leaving total expenditures relatively unchanged. The analysis therefore tests for effects on the total financial burden borne by households, which is arguably the most policy-relevant measure, but cannot rule out increases in underlying electricity rates that are offset by demand responses. The institutional structure of U.S. electricity pricing, with regulated rates set uniformly across utility service territories and separate tariff schedules for large industrial loads, is designed to limit precisely the kind of cost pass-through our analysis tests for. Aggregate, regional-wholesale-market-level effects of data center demand growth remain an open question.

The paper makes three contributions. We construct what is, to our knowledge, the most extensive dataset linking data center locations to household electricity expenditures, covering 24 states and a decade of data at both the ZIP code and utility territory levels. The dataset is built entirely from public sources (Census PUMS, interconnection queue filings, and EIA Form 861) and can serve as a foundation for future research. We also find that our data do not robustly support the commonly expressed concern about data centers "raising electricity bills" at either the local or utility-territory level during the pre-hyperscaler period, providing an empirical counterpoint to a widely repeated claim, though we emphasize that wide confidence intervals and measurement limitations mean our results constitute absence of evidence rather than evidence of absence. Finally, by conducting analysis at both the ZIP code level and the utility territory level, and showing why the latter is the more institutionally appropriate unit, we illustrate the value of aligning the unit of analysis with the institutional level at which the outcome is determined, a methodological point that applies beyond the data center context.

Several follow-up directions are discussed in Section 7.4. The most useful is obtaining *individual household-level* electricity billing data, through utility partnerships or state data-sharing programs, which would overcome the ecological inference limitations of our ZIP-code-level approach and enable matched household comparisons that control for home size, building age, and baseline consumption. Such data could detect effects that are averaged away in ZIP-level aggregation and could reveal distributional impacts across income groups and housing types. Qualitative research on household perceptions of data centers and electricity costs before the issue becomes fully politicized is also valuable: the window for collecting unbiased baseline data on how residents perceive, attribute, and respond to electricity cost changes is narrowing as media coverage intensifies and advocacy groups mobilize. As experience with wind turbines and natural gas pipelines shows, once an infrastructure siting issue becomes partisan, evidence-based policy becomes much harder.

As data centers continue their expansion across the United States, driven by the energy demands of artificial intelligence training and inference, their effects on the broader electricity system, on residential ratepayers, and on host communities will remain an open question at the intersection of energy economics, technology policy, and infrastructure planning. Decisions made in the next few years about data center siting, rate design, grid investment, and community compensation will shape the relationship between the digital economy and the electricity system for decades. Our analysis provides a first empirical foundation for that inquiry, finding that the most commonly voiced concern (localized bill increases) is not supported by the available evidence during the pre-hyperscaler period, while clearly delineating what current data can and cannot tell us. The measurement limitations we identify (the PUMA-to-ZIP spatial smoothing, incomplete data center inventories, and the bills-versus-rates distinction) point toward the more granular, multi-dimensional research agenda that this policy question demands.

References

Baker, Robert, "Data Centers Will Drive Up Electric Rates for Georgians," The Atlanta Journal-Constitution 2024. December 6, 2024; accessed May 2026.

Bertrand, Marianne, Esther Duflo, and Sendhil Mullainathan, "How Much Should We Trust Differences-in-Differences Estimates?," *Quarterly Journal of Economics*, 2004, *119* (1), 249–275.

Blunt, Katherine and Jennifer Hiller, "America's Biggest Power Grid Operator Has an AI Problem—Too Many Data Centers," The Wall Street Journal 2026. Accessed March 2026.

Borenstein, Severin, "The Redistributional Impact of Nonlinear Electricity Pricing," *American Economic Journal: Economic Policy*, 2012, *4* (3), 56–90.

_ **and James Bushnell**, "The U.S. Electricity Industry after 20 Years of Restructuring," *Annual Review of Economics*, 2015, *7*, 437–463.

Bushnell, James, Erin Mansur, and Kevin Novan, "Review of the Economics Literature on US Electricity Restructuring," *Working Paper*, 2017. University of California, Davis.

Callaway, Brantly and Pedro H. C. Sant'Anna, "Difference-in-Differences with Multiple Time Periods," *Journal of Econometrics*, 2021, *225* (2), 200–230.

Cameron, A. Colin and Douglas L. Miller, "A Practitioner's Guide to Cluster-Robust Inference," *Journal of Human Resources*, 2015, *50* (2), 317–372.

Davis, Lucas W., "The Effect of Power Plants on Local Housing Values and Rents," *Review of Economics and Statistics*, 2011, *93* (4), 1391–1402.

de Chaisemartin, Clément and Xavier D'Haultfœuille, "Two-Way Fixed Effects Estimators with Heterogeneous Treatment Effects," *American Economic Review*, 2020, *110* (9), 2964–2996.

de Vries, Alex, "The Growing Energy Footprint of Artificial Intelligence," *Joule*, 2023, *7* (10), 2191–2194.

Electric Power Research Institute, "Powering Intelligence: Analyzing Artificial Intelligence and Data Center Energy Consumption," Technical Report, EPRI 2024. Palo Alto, CA.

Georgia Public Service Commission, "Data Center Fact Sheet," 2026. March 2026.

Goodman-Bacon, Andrew, "Difference-in-Differences with Variation in Treatment Timing," *Journal of Econometrics*, 2021, *225* (2), 254–277.

Ho, Daniel E., Kosuke Imai, Gary King, and Elizabeth A. Stuart, "Matching as Nonparametric Preprocessing for Reducing Model Dependence in Parametric Causal Inference," *Political Analysis*, 2007, *15* (3), 199–236.

Imbens, Guido W. and Donald B. Rubin, *Causal Inference for Statistics, Social, and Biomedical Sciences: An Introduction*, New York: Cambridge University Press, 2015.

International Energy Agency, "Electricity 2024: Analysis and Forecast to 2026," Technical Report, International Energy Agency 2024. Paris, France.

Jarvis, Stephen, "The Economic Costs of NIMBYism: Evidence from Renewable Energy Projects," *The Economic Journal*, 2022, *132* (645), 1535–1573.

Joskow, Paul L., "Capacity Payments in Imperfect Electricity Markets: Need and Design," *Utilities Policy*, 2008, *16* (3), 159–170.

_ , "Lessons Learned from Electricity Market Liberalization," *The Energy Journal*, 2008, *29* (Special Issue 2), 9–42.

Koomey, Jonathan, "Growth in Data Center Electricity Use 2005 to 2010," *Analytics Press*, 2011. Oakland, CA.

Lee, Vivian, Ross LaFleur, Khushboo Goel, and Bassem Khoury, "The Impact of GenAI on Electricity: How GenAI Is Fueling the Data Center Boom in the U.S.," Boston Consulting Group 2023. September 13, 2023.

Loudoun County Department of Economic Development, "Data Centers in Loudoun County: Economic Impact Report," 2023. Leesburg, VA.

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

Miller, Rich, "Tax Breaks and Data Centers: Are Incentive Programs Worth It?," Data Center Frontier 2021. Accessed March 2026.

Missouri Census Data Center, "Geographic Correspondence Engine (Geocorr)," https://mcdc.missouri.edu/applications/geocorr.html 2022. Accessed March 2026.

Moulton, Brent R., "An Illustration of a Pitfall in Estimating the Effects of Aggregate Variables on Micro Units," *Review of Economics and Statistics*, 1990, *72* (2), 334–338.

Mytton, David, "Data Centre Water Consumption," *npj Clean Water*, 2021, *4*, 11.

National Renewable Energy Laboratory, "U.S. Electric Utility Companies and Rates: Look-up by ZIP Code (2019)," https://data.openei.org/submissions/4042 2019. OpenEI Utility Rate Database. Accessed 2026.

Oregon Public Utility Commission, "Large Customer Demand," 2026. Accessed May 2026.

Paullin, Charlie, "Dominion seeks to expand rate program designed to shift grid demand," Virginia Mercury 2023. Accessed March 2026.

Penn, Ivan and Karen Weise, "Big Tech's A.I. Data Centers Are Driving Up Electricity Bills for Everyone," The New York Times 2025. Accessed March 2026.

Rogoway, Mike, "Oregon's Data Centers Want a Lot More Electricity. Who's Going to Pay? It Could Be You," The Oregonian/OregonLive 2024. October 13, 2024; accessed March 2026.

Shehabi, Arman, Sarah Josephine Smith, Alex Hubbard, Alexander Newkirk, Nuoa Lei, Md AbuBakar Siddik, Billie Holecek, Jonathan G. Koomey, Eric R. Masanet, and Dale A. Sartor, "2024 United States Data Center Energy Usage Report," Technical Report LBNL-2001637, Lawrence Berkeley National Laboratory 2024. Berkeley, CA. Prepared for the U.S. Department of Energy.

_ **, Sarah Smith, Dale Sartor, Richard Brown, Magnus Herrlin, Jonathan Koomey, Eric Masanet, Nathaniel Horner, Inês Azevedo, and William Lintner**, "United States Data Center Energy Usage Report," Technical Report LBNL-1005775, Lawrence Berkeley National Laboratory 2016. Berkeley, CA.

Sun, Liyang and Sarah Abraham, "Estimating Dynamic Treatment Effects in Event Studies with Heterogeneous Treatment Effects," *Journal of Econometrics*, 2021, *225* (2), 175–199.

U.S. Department of Energy, "AI for Energy: Opportunities for a Modern Grid and Clean Energy Economy," Technical Report, U.S. Department of Energy 2024. Washington, DC.

U.S. Energy Information Administration, "Form EIA-861M (Monthly Electric Power Industry Report)," https://www.eia.gov/electricity/data/eia861m/ 2026. Accessed June 2026.

Virginia Joint Legislative Audit and Review Commission, "Data Centers in Virginia," Technical Report, Joint Legislative Audit and Review Commission 2024. Richmond, VA.

Virginia State Corporation Commission, "Proceeding on Dominion Energy Virginia's 2024 Rate Case," 2024. Case No. PUR-2024-00036, Richmond, VA.

A Data Sources

The panel uses four public sources. Collection scripts are available on request.

Data centers. Interconnection.fyi (`https://www.interconnection.fyi`) compiles state-level interconnection-queue filings from utilities and regional grid operators. We collect every record with "Operational" status for the 24 states in the sample. For records that list a city or street address but no ZIP code, we geocode to coordinates and assign the closest ZIP, constraining the candidate ZIPs to the listed state to prevent border misassignments. Records that cannot be geocoded to a valid in-state ZIP are kept in state totals but excluded from the ZIP-level analysis. A small share of project pages return errors when accessed (between 5% and 10% across states) and are dropped as stale listings.

PUMA-to-ZIP crosswalks. Geocorr (`https://mcdc.missouri.edu/applications/geocorr.html`) from the Missouri Census Data Center provides population-weighted PUMA-to-ZCTA crosswalks. We use the 2018 release (2012-vintage PUMAs) for ACS years 2014–2021 and the 2022 release (2020-vintage PUMAs) for ACS years 2022–2024, since the Census Bureau redrew PUMA boundaries after the 2020 census.

ACS PUMS. For each state and each year from 2014 through 2024 (excluding 2020), we use the variables `ELEP` (monthly electricity cost), `WGTP` (household weight), `TEN` (tenure status), `HINCP` (household income), and `PUMA` (Public Use Microdata Area). PUMS is a public-use sample of approximately 1% of the population per year. The 2020 ACS was not released in standard 1-year PUMS form because of pandemic-related collection problems, which is why the panel skips that year.

ZIP-to-utility mapping. A 2019 snapshot of EIA Form 861, redistributed through NREL Open Energy Information. We prefer investor-owned utility (IOU) coverage and use customer count to break ties when a ZIP is served by multiple IOUs; non-IOU coverage fills in the remaining ZIPs.

B State Coverage

Table 13 provides detailed statistics for each state in the sample.

Table 13: Detailed State-Level Statistics

State	ZIP Codes	Data Centers	Avg. Bill ($)	Market Type
Arizona	417	72	157.99	Regulated
California	1,808	438	159.52	Restructured
Colorado	530	33	137.69	Regulated
Connecticut	289	0	185.35	Restructured
Florida	1,013	6	178.45	Regulated
Georgia	756	27	193.56	Regulated
Illinois	1,399	51	157.71	Restructured
Indiana	808	5	171.65	Regulated
Iowa	972	10	159.64	Regulated
Michigan	994	5	146.63	Hybrid
Minnesota	896	10	154.38	Regulated
Missouri	1,045	13	170.94	Regulated
Nevada	182	5	151.25	Regulated
New Jersey	599	50	170.98	Restructured
New York	1,819	18	160.54	Restructured
North Carolina	853	8	176.59	Regulated
Ohio	1,237	18	152.15	Restructured
Oregon	430	268	139.07	Regulated
Pennsylvania	1,848	11	152.89	Restructured
Tennessee	638	7	178.79	Regulated
Texas	1,995	651	187.43	Restructured
Virginia	915	552	177.75	Regulated
Washington	604	15	146.51	Regulated
Wisconsin	787	4	145.59	Regulated
Total	**22,834**	**2,277**	**163.83**	

Counts are state-ZIP units. Facilities without valid ZIPs, ZIP 00000, or post-2024 operational dates are excluded.

C Data Center Construction Timing by State

Figure 13 shows the distribution of data center operational years for each state. Blue bars indicate facilities that became operational before the panel period (pre-2014); red bars indicate within-panel construction (2014–2024). The red dashed line marks the start of the panel period. States are ordered by total data center count (highest to lowest). The timing of data center construction varies substantially across states: Virginia and Oregon have large stocks of pre-existing facilities, while Texas shows more recent, rapid growth concentrated in 2019–2023. Several states (WI, MI, NV) have only a handful of facilities, limiting within-state variation for causal identification.

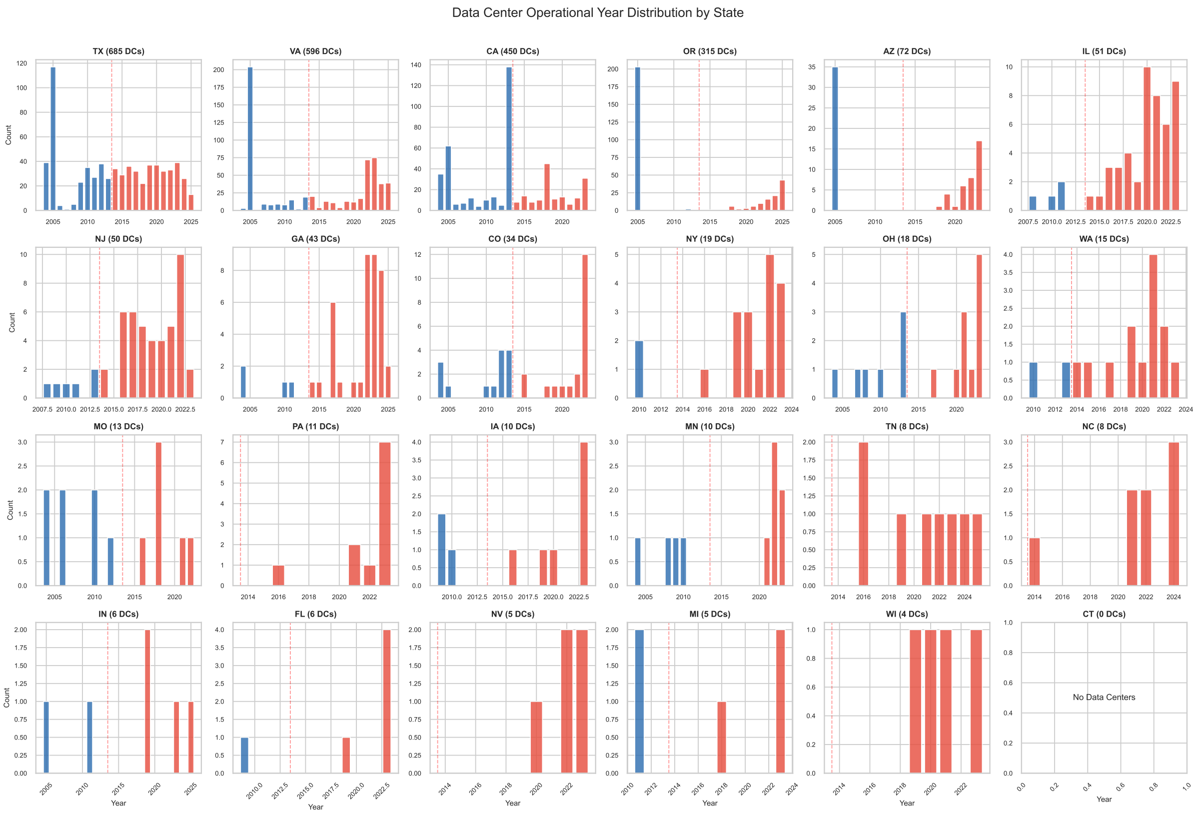

Figure 13: Data center operational year distribution by state. Blue: pre-panel (before 2014). Red: within panel (2014–2024). Red dashed line marks 2014 (panel start). States ordered by total DC count. Subplot labels display counts from the full scraped database (n = 2,434), which includes facilities excluded from the analytical sample of 2,277 because of post-2024 operational dates, invalid ZIPs, or missing geocoding.

D State-by-State Figures

This appendix presents state-by-state versions of the exploratory analyses shown in the main text. Each grid arranges the 24 states in order of average electricity bill (highest to lowest), with the number of operational data centers shown in parentheses.

D.1 Bill Trends by State

Figure 14 shows the average monthly electricity bill trajectory for each of the 24 states in the sample. All states share the common pattern identified in Section 5: relative stability from 2014–2019 followed by a sharp increase from 2021–2024. The magnitude of the post-2020 increase varies, with some states (e.g., GA, TX) experiencing larger jumps than others (e.g., WA, OR).

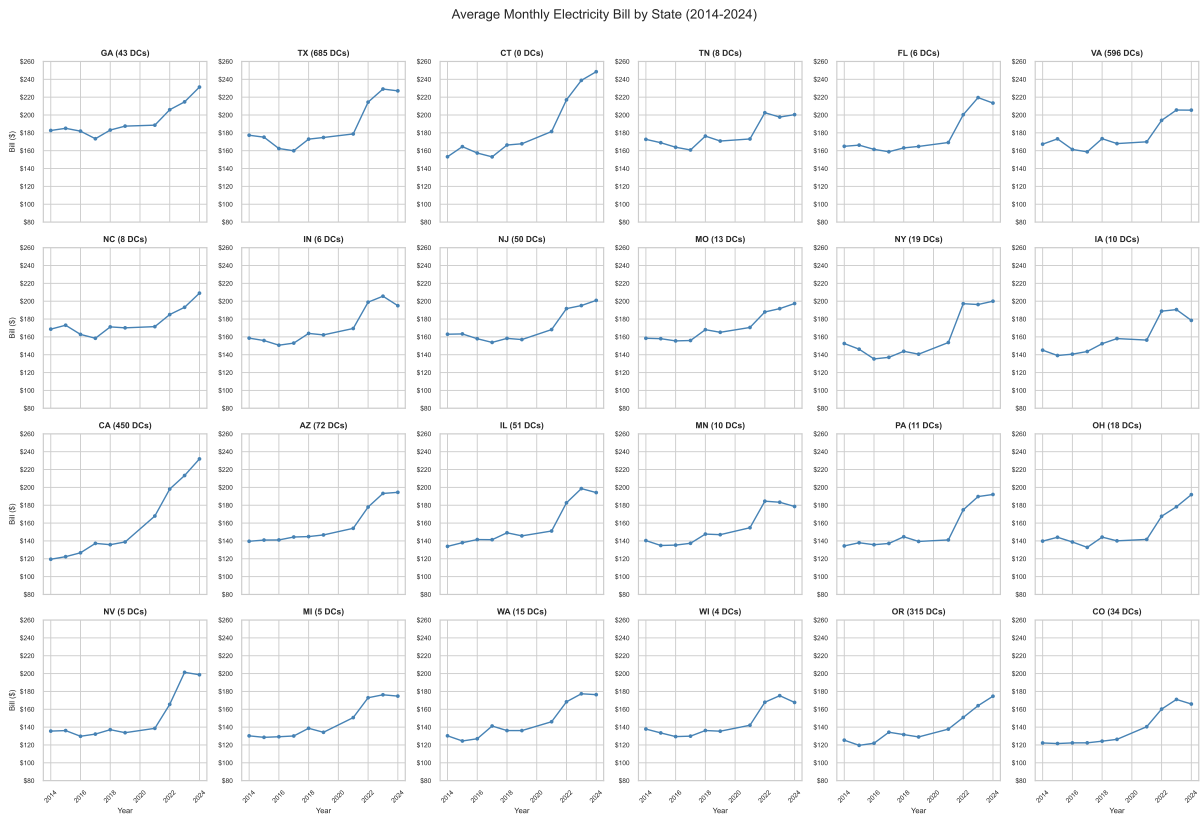

Figure 14: Average monthly electricity bill by year for each of the 24 states. States ordered by average bill level (highest to lowest). Data center count shown in parentheses.

D.2 Year-over-Year Bill Changes by State

Figure 15 displays year-over-year percentage changes for each state. The 2022 spike is visible in nearly every state, confirming it as a national phenomenon rather than a state-specific event. Some states show idiosyncratic patterns: Connecticut experienced a particularly large 2022 increase, while Oregon's changes were more moderate throughout.

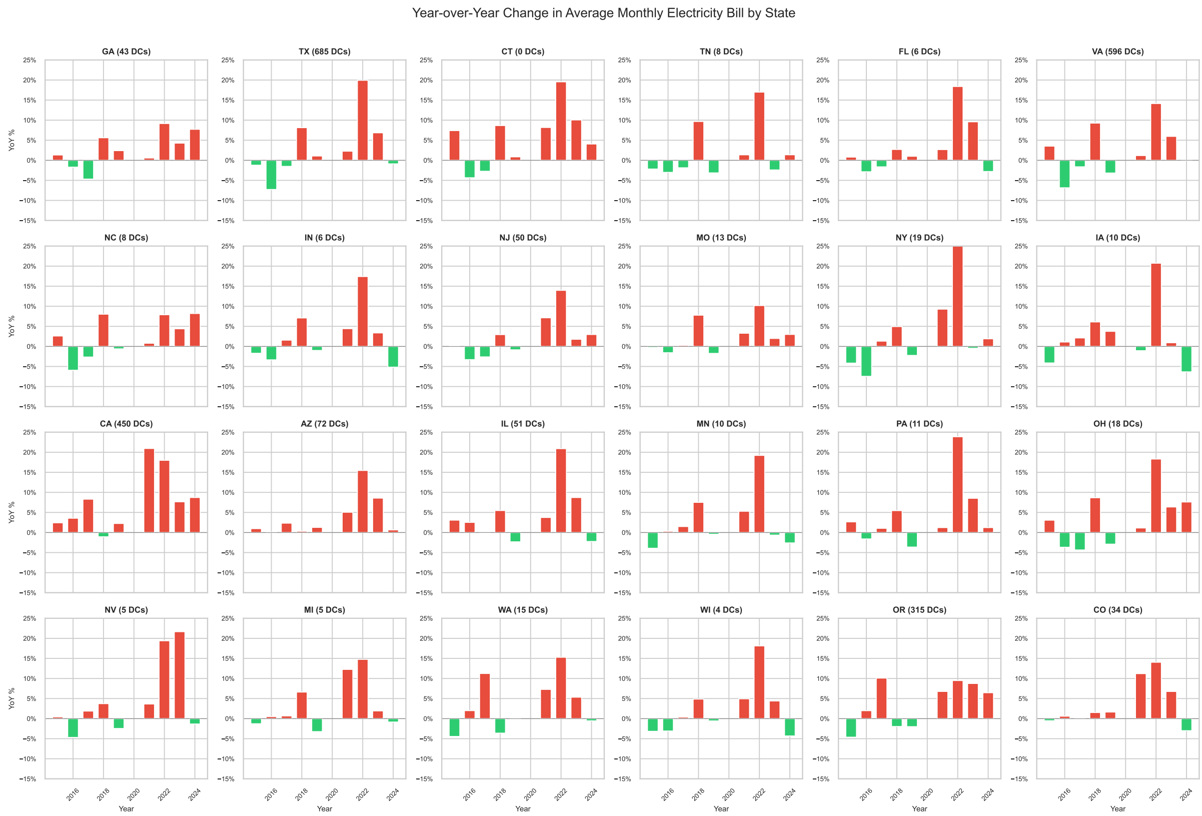

Figure 15: Year-over-year percentage change in average monthly electricity bill for each state. Red bars indicate increases; green bars indicate decreases.

D.3 DC ZIPs vs. Non-DC ZIPs by State

Figure 16 compares bill trajectories for data center ZIP codes (red solid) and non-DC ZIP codes (blue dashed) within each state. In most states, the two series track each other closely, with no visible divergence during the period of data center expansion. In some states (e.g., TX, OR), DC ZIPs actually have *lower* average bills than non-DC ZIPs. Connecticut, with zero data centers, serves as a pure control state. The state-by-state evidence corroborates the null finding from the causal analysis: there is no consistent pattern of DC ZIPs experiencing faster bill growth than non-DC ZIPs.

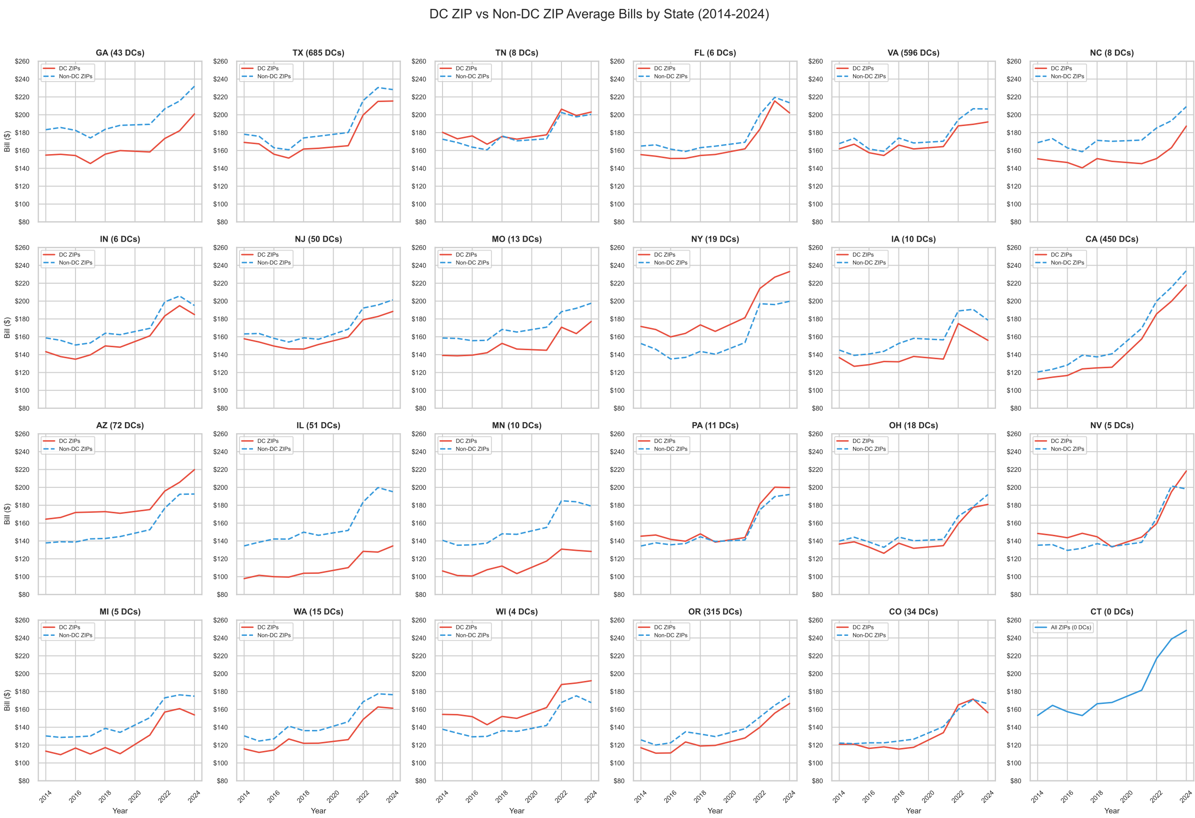

Figure 16: Average monthly electricity bills for data center ZIPs (red solid) versus non-DC ZIPs (blue dashed) within each state. Connecticut has zero data centers and shows only the all-ZIP average.

E Cohort DiD Event-Time Aggregation

Table 14 reports a cohort-size-weighted difference-in-differences diagnostic at each event time $e = t - g$, where g is the first-DC-entry year for a clean-treated cohort. Never-treated state-ZIPs serve as the comparison group; the baseline period for each cohort is $g - 1$. This is not a fully formal Callaway–Sant'Anna estimator with inference. Long-lead and long-lag cells are thinly populated and should be interpreted with caution; the noisiest leads are at $e = -10$ and the noisiest lags at $e \geq 8$.

Table 14: Cohort DiD ATT(e) Event-Time Aggregation (ZIP Level)

e	ATT(e) (\$/mo)	e	ATT(e) (\$/mo)
-10	7.99	0	-0.09
-9	-1.02	1	1.21
-8	1.38	2	0.97
-7	0.96	3	2.09
-6	1.40	4	5.97
-5	1.24	5	4.68
-4	-0.09	6	7.45
-3	-0.13	7	0.33
-2	0.33	8	4.30
		9	14.77
Average post-treatment ($e \geq 0$): \$2.44/mo			

Note: Aggregation uses cohort-size weights (number of clean-treated state-ZIPs in each cohort). $e = -1$ is the baseline (omitted by construction). The leftmost cells at $e \leq -7$ are based on only a few cohorts.