

Data Centers and Retail Electricity Prices: Evidence from County Openings, 2001–2024

Claude

October 2026

Claude is an AI system developed by Anthropic; this paper was written by the Claude Fable 5 model generation. This paper was commissioned by Jason Furman. The county retail prices were built from the Energy Information Administration's Form EIA-861 files, the facility list from the Environmental Protection Agency's ICIS-Air and Facility Registry Service data, and the establishment counts from the Census Bureau's County Business Patterns; the Global Data Center Map (Ringmast4r) was used as a facility universe under its attribution license (Data centers © Ringmast4r – Global-Data-Center-Map, <https://github.com/Ringmast4r/Global-Data-Center-Map>). The views expressed are those of the author. Replication code and the data manifest accompany the paper.

Data Centers and Retail Electricity Prices: Evidence from County Openings, 2001–2024

Claude · October 2026

ABSTRACT

Does a data center raise the retail electricity prices of the households and businesses around it? We build county-level retail prices by customer class for 2001 to 2024 from utility filings and service territories, and a list of 623 data-center facilities dated from the Environmental Protection Agency’s Clean Air Act records, and estimate the effect of a county’s first data center with staggered-adoption event studies. The first opening moved the residential price by -0.4 log points over six years (standard error 0.7): a few dollars a month either way under a Rambachan–Roth restriction that allows post-opening violations of parallel trends half the size of the largest pre-opening one. The serving utilities’ sales did not detectably move, so the typical first facility was a small load, and under a cost-of-service model in which the pass-through is the load share times an incremental-cost premium, the design bounds that premium only at an assumed load. Splits by facility class and utility ownership carry the model’s signs but do not survive inference fit to their size; a utility-level design in vertically integrated states shows a residential rise the county design does not. The far larger loads now being connected are the untested case.

JEL Codes: L94, L51, Q41, L86, R11

Claude
c/o Jason Furman
Harvard Kennedy School
79 JFK Street, Cambridge, MA 02138

1 Introduction

Loudoun County, Virginia, holds the largest concentration of data centers in the world, and the question of who pays for them is now a regulatory one: in 2024 the Virginia legislature’s audit arm asked whether the households and businesses already on Virginia’s utilities’ systems were paying for the new load (Joint Legislative Audit and Review Commission, 2024; E3, 2026), and in 2025 the State Corporation Commission’s order in Dominion’s biennial review created a separate rate class for customers of 25 megawatts or more (Virginia State Corporation Commission, 2025). The same question is now before commissions in Georgia, Ohio, Arizona, and a dozen other states. This paper asks it of the record so far: when a county received its first data center, did the retail electricity prices paid by its residential, commercial, and industrial customers move, relative to counties that had not yet received one, and by how much?

Theory does not settle the sign. Under cost-of-service regulation a utility recovers its embedded cost through rates, so a new load that pays the class tariff lowers everyone else’s price when the incremental cost of serving it is below the average cost it pays, which is the load-growth dividend of the 1960s and of the flat-demand decade after 2008, and raises it when the incremental cost is above, as when the load requires new generation or transmission, or is served under a discount. Which regime describes the data-center wave is a question about one gap, the incremental cost of the load against the tariff it pays, and the answer can differ by customer class, by the regime the utility operates under, and by how large the load is relative to the system that serves it.

We answer it with two constructions and a model. The first is a county-level panel of retail prices by customer class from 2001 to 2024, built from the Energy Information Administration’s utility-by-state filings and the service territories each utility reports, following the Administration’s own arithmetic; it reproduces the existing county series for 2012 onward exactly and extends it back eleven years. The second is a list of 623 data-center facilities with opening dates drawn from the Environmental Protection Agency’s Clean Air Act records, where a data center appears when it permits its backup generators; the list carries a stated date quality for every facility and was tested against 31 documented hyperscale openings. The treatment is the year of a county’s first facility, 148 counties in all, and the design is a staggered-adoption event study estimated three ways that are robust to heterogeneous effects, Sun–Abraham, Callaway–Sant’Anna, and a stacked design, against the two-way fixed-effects benchmark, with

the counties that never receive a facility as controls. The model is a cost-of-service utility with an embedded-cost allocator and one large new load, in which the pass-through to incumbents is the load's share of sales times its incremental-cost premium, and which is calibrated to the data.

The first finding is a bound. The first data center moved the residential price of the county that received it by -0.4 log points over the six years after the opening, with a standard error of 0.7 . The pre-opening coefficients reject parallel trends at the margin ($p = 0.05$ under the headline estimator), so the bound the paper stands on is the one that allows post-opening violations half the size of the largest pre-opening one: -5.0 to 4.5 log points, which at the treated counties' pre-opening price and consumption is about \$5 a month either way on a bill of about \$1,280 a year. At the full size of the largest pre-opening divergence the set is -9.1 to 8.7 log points, about twice as wide. Under parallel trends the interval is -1.8 to 1.0 log points, between a saving of \$22 and an increase of \$12 a year. The three robust estimators agree with each other and with the benchmark, and a never-treated control group matched to the treated counties on their pre-period price level and growth gives -0.0 (0.7).

The second finding is that the typical first opening was a small load, which is why the estimate bounds the effect without deciding its sign. The model says the effect scales with the load's share of the serving utility's sales, and the paper's own first stage, the same event study run on the serving utilities' commercial and industrial sales, finds no detectable change at the first opening: -1.8 log points (1.3). Facility load is not in the public record, and a proxy that sets every facility at 100 MW against its serving utility's sales varies only with the utility's size; on it the residential effect moves by -0.02 log points per percentage point of share (0.08), a slope indistinguishable from zero. At the headline precision the design can detect a change of 2.0 log points, which at a 100 MW load and the median share is a premium of 84 percent of the retail price, and at a 10 MW load ten times that. The premium of the typical opening is therefore bounded only on an assumption about its size.

The third finding is that the cells where a sign appears do not survive inference that fits their size. Where the first facility was a Clean Air Act major source, a large bank of backup generators, the residential price fell by 1.9 log points on the state-clustered standard error (0.7), but a wild-cluster bootstrap on the 15 states that carry treatment gives $p = 0.41$. In the 8 counties whose largest utility is a municipal or public system, the residential price rose by 4.5 log points (2.5) and every other class rose with it, on 3 treated states and a bootstrap p of 0.22 . The signs are

the ones the model predicts; the paper reports them as signs and not as findings. The model has a prediction for the second cell: a system that buys its power has no embedded energy cost to average a cheap megawatt-hour into, so it loses the dividend a vertically integrated utility passes on. At the utility unit, where rates are set, a vertically integrated utility whose largest county received a first facility shows a residential rise of 2.8 log points (1.1) against a pool of never-treated utilities two orders of magnitude smaller, and 1.1 (1.7) against the 6 never-treated utilities of comparable size; the county design and the utility design disagree on the residential sign in that regime, and the paper reports the disagreement as open.

The mechanism is the arithmetic of embedded cost. The fixed cost of the system is about half of the retail price, so a new customer paying the system-average tariff contributes half its bill to costs it did not cause, and under a sales allocator the incumbents' price falls unless the load's incremental energy and network cost exceeds the average by that much. A customer billed on the commercial tariff pays about the system average toward fixed cost on the paper's class prices, 1.01 of it per kilowatt-hour, but one billed on the industrial tariff pays 0.47 of it, which shrinks the dividend by about half; a discounted contract, a large uncontributed interconnection, or a purchased-power system in which the load is a large share of sales reverse the sign; the large-load tariffs now spreading through state commissions, with minimum demand charges and contributions in aid of construction, move it back. The paper's reading of its estimates is that the openings of 2002 to 2024 did not raise the residential price of the typical county by more than a few dollars a month, that the data cannot say whether the typical opening paid its own incremental cost, and that the places where prices rose are the ones the model points to: small systems that buy their power.

This paper relates to three literatures, which Section 2 takes strand by strand. The first is the direct evidence on data centers and retail prices, all of it from 2026 and all preliminary (Watten et al., 2026; Garimella, 2026; Scalera et al., 2026; Alvarez et al., 2026; Kay et al., 2026; Golden et al., 2026); relative to it, this paper uses the county as the unit over 24 years, a public and dated treatment with a stated error structure, outcomes by customer class, a first stage on the serving utilities' sales, and a model that says what a bound means. The second is the economics of cost-of-service regulation and rate design (Joskow and Schmalensee, 2024; Lazar et al., 2020; Borenstein, 2016; Borenstein and Bushnell, 2022), from which the model takes the allocator, the contribution in aid of construction, and the lag and makes them the parameters the data identify.

The third is the econometrics of staggered adoption (Goodman-Bacon, 2021; Callaway and Sant’Anna, 2021; Sun and Abraham, 2021; Roth et al., 2023; Cengiz et al., 2019; Rambachan and Roth, 2023), which the paper applies as that literature prescribes.

Two caveats belong on the first page. The facility list is built from a federal register that some states do not fill and whose dates are late by a year or more in most documented cases and early in one, so the county estimates are attenuated in a direction the known-cases test establishes and by an amount it cannot. And the wave of facilities dated 2025 and later, 27 counties with no post-opening year in the data, is an order of magnitude larger than the openings studied here; the paper says what the model predicts at those load shares and does not claim to have measured them.

Section 2 places the paper in the literature. Section 3 describes the rate case, regulatory lag, restructuring, and the large-load tariffs. Section 4 sets out the model and its four propositions. Section 5 describes the two constructions and the sample. Section 6 states the estimand, the design, the threats, and the diagnostics. Section 7 reports the estimates and their magnitude. Section 8 reports the battery and the utility-level reconciliation. Section 9 reports the dose-response and the splits, with the inference their size allows. Section 10 reads the estimates through the calibrated model and in dollars. Section 11 concludes.

2 Related literature

Three strands of work bear on the question.

The direct evidence on data centers and retail prices is a year old. Watten et al. (2026) assemble a state-level panel of data-center capacity and retail prices for 2015 to 2024 and instrument capacity with the mileage of the 1947 interstate highway plan; they find that more capacity lowered average residential prices, and they find in Form 1 cost data suggestive evidence of the mechanism, economies of scale in transmission, distribution, and generation cost per kilowatt-hour. Their analytic appendix states the condition this paper’s model makes its first proposition: prices fall when the marginal cost of the new load plus the shadow cost of capacity is below the average cost. Garimella (2026) builds a ZIP-level panel in 24 states for 2014 to 2024 with facility opening dates scraped from interconnection-queue filings and household bills from the American Community Survey; a two-way fixed-effects difference-in-differences gives a small and imprecise increase in monthly bills, a matched version smaller still, and the

household-weighted utility-territory version a decrease, and the author reads the evidence as not supporting a material bill effect in the pre-hyperscaler era while noting that the facility list likely misses operating facilities, which biases the estimate toward zero. Scalera et al. (2026), whose discussion paper the authors could read only in abstract, map the post-2021 build-out to utility territories and report that a standard deviation of ten facilities raised household prices by 13 percent and a single hyperscale facility by 4 percent, that households bear a disproportionate share relative to the other classes, and that nonprofit ownership, energy surplus, and retail competition may mitigate the effect. Alvarez et al. (2026) assemble a facility-level list of U.S. data centers by county and, with a shift-share design on fiber-backbone geography, find positive effects on employment, establishments, income, house prices, water withdrawals, and electricity prices, the one county-level price result in the literature, obtained as one of many outcomes. On the wholesale side, Kay et al. (2026) model hourly dispatch and attribute a few percent of wholesale prices to existing data centers, with far more in forward scenarios; Golden et al. (2026) find in a difference-in-differences on model releases that AI demand raised fossil generation and wholesale prices in treated zones; Knittel et al. (2025) show that flexible data centers lower system cost; and Muller (2026) price the air-pollution and greenhouse-gas damages of the existing fleet.

The reports that frame the policy debate are the Lawrence Berkeley retail-price studies, Wiser et al. (2025, 2026), who find that states with more load growth since 2019 saw smaller retail price increases through fixed-cost spreading, a relation that is weaker for residential prices alone, and Forrester et al. (2024), who state the caveat the three reports share, that load growth can raise or lower prices depending on the cost of new supply, tariff structure, and, as the 2024 report adds, regulatory lag; E3 (2026), who find no quantitative evidence of a historical subsidy and count the large-load tariffs; Martin and Peskoe (2025), who document the discounts and confidentiality in nearly fifty proceedings; and the consumption accounting of Shehabi et al. (2016, 2024), Electric Power Research Institute (2024), and International Energy Agency (2025). What the strand lacks, and this paper adds, is a county retail panel that predates 2014, a facility list that is public and dated with a stated error, outcomes by customer class, and a model that says what a null means.

The second strand is the economics of cost-of-service regulation. The revenue requirement, the class cost-of-service study, and the allocators that turn one into tariffs are set out

in Joskow and Schmalensee (2024), Lazar et al. (2020), and Lazar and Gonzalez (2015), and the public-utility pricing problem more generally in Wolak (2008); regulatory lag, the fact that rates are constant between cases, is in Joskow (1974) as reported by Joskow and Schmalensee (2024), and its incentive properties in Joskow and Schmalensee (1986) and Joskow (2008). The recovery of fixed cost through volumetric rates, and the gap between retail prices and marginal cost that follows, is the subject of Borenstein (2016) and Borenstein and Bushnell (2022), whose national figures put private marginal cost at roughly a third of the average residential price; Watten et al. (2026) build on them to put marginal cost at about half of average cost, which is the paper's calibration of the embedded fixed-cost share. Restructuring and what it did to cost and prices is in Borenstein and Bushnell (2015), Fabrizio et al. (2007), Cicala (2015), and Cicala (2022); the pass-through of cost shocks in electricity markets in Fabra and Reguant (2014) and of energy cost in manufacturing in Ganapati et al. (2020); the local incidence of a generating plant in Davis (2011); and the rate-design agenda for distributed resources in MIT Energy Initiative (2016). This paper takes the allocator, the contribution in aid of construction, and the lag from this strand and makes them the parameters of a comparative static that a staggered design can test.

The third strand is the econometrics the design uses. Goodman-Bacon (2021) and de Chaisemartin and D'Haultfœuille (2020) showed what the two-way fixed-effects estimator averages under staggered adoption; Callaway and Sant'Anna (2021), Sun and Abraham (2021), Borusyak et al. (2024), and Gardner (2022) give estimators that do not use already-treated units as controls, surveyed in de Chaisemartin and D'Haultfœuille (2023) and Roth et al. (2023), and Baker et al. (2022) show how much the choice matters in practice; Cengiz et al. (2019) is the stacked clean-control design the paper's dose-response is estimated on. On pre-trends, Roth (2022) shows that passing a pre-test is weak evidence, Freyaldenhoven et al. (2019) propose a correction, and Rambachan and Roth (2023) give the confidence sets under bounded violations the paper reports. On inference with clustered errors, Bertrand et al. (2004), Cameron et al. (2008), and MacKinnon and Webb (2018) are the reasons for state clusters and the wild bootstrap; on synthetic controls, Abadie et al. (2010) and Abadie (2021) are the method and its feasibility conditions for the Virginia exercise. This paper claims no contribution to the strand.

Against the three closest papers the differences are these. Watten et al. (2026) work at the state level with an instrument for capacity and find prices fall; this paper works at the county level with openings as events and finds a bound that includes their sign at the average and their sign

where the dose is large. Garimella (2026) works at the ZIP level over 2014 to 2024 with scraped dates and finds no effect, conceding attenuation; this paper’s list is public, its dates have a measured error, and its gradient tests the model where attenuation cannot explain the sign. Scalera et al. (2026) work at the utility-territory level over the post-2021 wave and find prices rise, by 4 percent per hyperscale facility; this paper’s period ends in 2024, when that wave had no post-opening years, and its split by hyperscale presence, 0.1 log points (0.8) on the earlier and smaller facilities, is the nearest comparison and does not reproduce it. Two of their findings read differently through this paper’s model: that households bear more than the other classes is what Proposition 3 predicts for the measured average of the class that bills the load, whatever the incumbents in that class pay; and that nonprofit ownership may mitigate the effect runs against this paper’s one cell of rising prices, the small public systems, for which the model’s purchased-power case predicts the rise. Alvarez et al. (2026) find a positive but imprecisely estimated county price effect with a fiber shift-share that identifies from a different source of variation than openings; this paper’s design cannot reject an effect of their size at the typical opening. The question of the larger wave is open and is a question about the load share and the tariff, which the model states and the next vintage of these data can test. Two literatures that the facility list touches are outside the question and are not engaged: the local economic effects of siting, employment, establishments, and house prices, which Alvarez et al. (2026) estimate, and the emissions and air-quality damages of the fleet, which Muller (2026) price.

3 Institutional setting

Retail electricity prices in the United States are set in a regulatory proceeding, and the question this paper asks is a question about that proceeding. Four features of it matter for the design and the model.

The first is the rate case. A utility that sells at regulated rates files a case with its state commission in which it establishes a revenue requirement, the sum of operating cost, depreciation, and a return on its rate base, and a class cost-of-service study allocates that requirement to residential, commercial, and industrial customers (Joskow and Schmalensee, 2024; Lazar et al., 2020). The study assigns energy-related cost by sales, capacity-related cost by a measure of each class’s contribution to peak demand, and customer-related cost by customer counts, and the tariffs that come out of it recover each class’s allocated share (Lazar and

Gonzalez, 2015). A customer with a high load factor, such as a data center that draws nearly the same power every hour of the year, is allocated less capacity cost per kilowatt-hour than a class whose demand is peaky, which is the allocation-intensity parameter the model carries.

Cost-of-service regulation and its incentive properties are the subject of Joskow and Schmalensee (1986) and Joskow (2008); what the design uses is only that rates are set to recover embedded cost, class by class.

The second is regulatory lag. Between cases, rates are constant apart from automatic adjustment clauses that pass through fuel and purchased-power cost. Joskow and Schmalensee (2024) record that a third of utilities had no formal rate review over 1958 to 1972; cases remain infrequent, and some states have moved to formula rates and multi-year plans; the calibration takes a case every three years as a round number. A new load's effect on rates therefore arrives at the next case, not at the opening, and the model's regulatory-lag profile is the event-study shape that follows.

The third is restructuring. In the states that restructured their retail markets in the late 1990s and early 2000s, the incumbent utility divested or ring-fenced its generation, customers may buy energy from competitive suppliers, and the utility's regulated business is the wires: transmission and distribution at cost of service, with a default-service energy charge procured in the market for customers who do not shop (Borenstein and Bushnell, 2015). The lists in American Public Power Association (2019) and National Renewable Energy Laboratory (2017) identify the states; the two sources disagree on California, where retail choice was suspended in 2001, on Michigan, where choice is capped, and on Montana; the paper's regime variable follows the APPA list, under which all three are restructured, and the three states hold few treated counties. In a restructured state a large new load does not change an embedded energy cost; it moves a wholesale price, and the sign of that effect on incumbents is different, which the model's restructured case and the regime split in Section 9 take up. Restructuring also changes what the data measure for the industrial class: an investor-owned utility in a restructured state reports the delivery-only sales of customers who shop beside the bundled sales of those who do not, so its industrial average revenue per kilowatt-hour mixes a wires charge with a full-service price in proportions that move as customers switch. Section 8 returns to this when it reconciles the utility-level and county-level estimates.

The fourth is the special treatment of large loads. The incremental cost of connecting a

facility of a few hundred megawatts, the discount a state may offer to attract it, and the risk that the facility leaves before its cost is recovered have produced a wave of large-load tariffs. E3 (2026) count at least 38 such tariffs established between 2018 and 2026, most of them in 2025 and 2026, with minimum demand charges, contributions in aid of construction, and exit fees. Martin and Peskoe (2025) review nearly 50 regulatory proceedings and document discounted special contracts, confidentiality provisions, and allocation mechanics that can shift a data center's cost to captive ratepayers. In the model these are the allocation intensity and the up-front contribution; in the data they are, for the most part, after the sample, which is a limitation the paper states where the estimates are read.

What the data measure is set by Form EIA-861, the Energy Information Administration's annual census of electric utilities. Each utility reports, for each state it serves, its revenue, sales, and customer count by class, so that revenue divided by sales is the class's average price per kilowatt-hour; since 2012 small utilities file a short form with totals only. A utility also lists the counties in its service territory. The county price used here is the sales-weighted average of the state-wide class prices of the utilities that list the county, which Section 5 describes. It is an average revenue, not a tariff, so it moves with the mix of customers and their consumption as well as with rates; the composition proposition in the model is about that distinction.

4 A model of cost recovery and a large new load

A regulated utility recovers its costs through rates set at periodic rate cases, and a large new customer changes both the costs to be recovered and the sales over which they are spread. The model in this section is the simplest one in which that arithmetic produces an estimating equation and a sign for its coefficient. It has one price-setting utility, three incumbent customer classes, an embedded-cost allocator, and one new load. Its purpose is to say, in the parameters a rate analyst would recognize, when the arrival of a data center raises the retail price paid by the customers already there, when it lowers it, and by how much. The general allocator and the proofs are in Appendix A.

4.1 *Environment*

Time is a single rate-case cycle: a before period, in which rates were set without the new load, and an after period, in which the regulator re-sets them with it. A single utility serves

incumbent customer classes $j \in \{\text{residential, commercial, industrial}\}$ with annual sales q_j megawatt-hours and total sales $Q = \sum_j q_j$. Serving the incumbents costs an embedded (average) energy cost c per megawatt-hour plus a fixed cost K per year, which stands for generation capacity, transmission, distribution, and administration. Rates must recover the revenue requirement $cQ + K$. The utility is vertically integrated and regulated on cost of service in the baseline; the restructured case, in which the energy charge is a market price, is treated below.

Rate design allocates the fixed cost to classes by an embedded-cost allocator. In the baseline the allocator is proportional to sales, so every incumbent class pays the same price per megawatt-hour,

$$p_j = c + k \quad (1)$$

where $k = K/Q$ is the embedded fixed cost per megawatt-hour of sales. A demand or customer allocator changes the incumbent classes' shares of K without changing the comparative statics that follow, which are stated for a class whose allocated share equals its sales share; Appendix A works through the general case.

A new large load L megawatt-hours per year arrives: a data center. Serving it costs the marginal energy cost m per megawatt-hour, which exceeds c when the system must add supply at above-embedded cost and falls short of c when there is slack, plus an annualized incremental network and capacity cost γ per megawatt-hour, of which a share ϕ is paid up front by the customer as a contribution in aid of construction and the remainder $1 - \phi$ enters the rate base. The new load's share of sales is $\lambda = L/(Q + L)$.

4.2 *The regulator's problem and the price after the new load*

At the next rate case the regulator averages the energy cost over all sales and adds the new load's uncontributed network cost to the fixed cost:

$$c' = \frac{Lm + Qc}{L + Q} \quad (2)$$

$$K' = L\gamma(1 - \phi) + Qk \quad (3)$$

The new load is allocated the share $\theta\lambda$ of the new fixed cost K' , where θ is the allocation intensity: $\theta = 1$ under a sales allocator; $\theta < 1$ under a demand allocator when the new load's load factor exceeds the system's, and under a discounted economic-development tariff; $\theta > 1$

when the regulator loads the new customer with more than its sales share, as a large-load tariff with minimum demand charges does. The incumbent classes jointly bear the remainder in their pre-existing proportions, so the price an incumbent class pays after the rate case is

$$p'_j = \frac{K'(-\lambda\theta + 1)}{Q} + c' \quad (4)$$

The change in the incumbent price, written in the new load's share λ , is

$$\Delta p_j = \lambda(-c + m) + \gamma(1 - \phi)(-\lambda\theta + 1) \frac{\lambda}{1 - \lambda} - k\lambda\theta \quad (5)$$

which Appendix A derives from equations (1) to (4). The first term is the energy channel: averaging the new load's marginal energy cost into the embedded rate. The second is the fixed cost the new load takes over from the incumbents. The third is the incumbents' share of the new load's uncontributed network cost. Equation (5) is the object the empirical sections estimate.

4.3 *The sign of the pass-through*

Under a sales allocator the three terms collapse to one product. Define the incremental-cost premium of the new load as its full incremental cost per megawatt-hour, net of what it contributes up front, minus the average cost the incumbents pay:

$$\pi = -c - \gamma(\phi - 1) - k + m \quad (6)$$

Proposition 1 (Sign and dose). *Under a sales allocator ($\theta = 1$), the change in every incumbent class's price is the new load's share of sales times its incremental-cost premium,*

$$\Delta p_j|_{\theta=1} = \lambda(-c - \gamma\phi + \gamma - k + m) \quad (7)$$

so the incumbent price rises if and only if $\pi > 0$, and the effect per unit of load share is $\partial\Delta p_j/\partial\lambda = \pi$.

Intuitively, the new load pays the average price $c + k$ on every megawatt-hour it buys, and it costs the system $m + (1 - \phi)\gamma$ on every megawatt-hour it takes. If the second exceeds the first the incumbents make up the difference in proportion to the load's share of sales; if the first exceeds the second the incumbents receive the difference. A load that imposes no new fixed cost at all ($\gamma = 0$) and costs the embedded energy rate to serve ($m = c$) has a premium of $-k$: it pays k per megawatt-hour toward fixed costs it does not cause, and every other class's price falls by λk .

In the data's units, the embedded fixed cost is about half of the retail price in the United States (Watten et al., 2026), so a 100 MW data center that is 3.5 percent of its serving utility's sales, the median in our data, and that costs the embedded energy rate to serve, lowers the price of every other class by about $\lambda k/p$, which is 3.5 percent times 50 percent before the calibrated network cost, or 1.4 log points net of it. Under a sales allocator, cost shifting onto households requires the incremental cost of the new load to exceed the average cost by more than the fixed cost it takes over, and the load-growth dividend is the default. Under the class tariffs actually charged the default depends on the class that bills the load: a load billed on the commercial tariff pays about the system-average fixed cost per kilowatt-hour, $\theta_C = 1.01$ of it on the paper's class prices, but one billed on the industrial tariff pays $\theta_I = 0.47$ of it, and at that allocation the same neutral load lowers the incumbents' price by 0.5 log points, about half the sales-allocator figure. The industrial class's lower price is partly cost-based, since a high-load-factor customer at transmission voltage causes less capacity and distribution cost, so θ_I read from the price ratio is an upper bound on the discount and the industrial-tariff rows an upper bound on the dividend lost; a load that causes less network cost also has a lower γ , which the rows hold fixed. Which case describes the data-center wave is what the event study measures.

4.4 Allocation, contributions, and the restructured regime

The sign in Proposition 1 is conditional on the sales allocator. Three parameters of rate design move it.

Proposition 2 (Allocation and regime). *For any $\lambda \in (0, 1)$ and $\theta\lambda < 1$: (i) the pass-through to incumbents falls in the allocation intensity θ , $\partial\Delta p_j/\partial\theta < 0$; (ii) it falls in the up-front contribution ϕ , $\partial\Delta p_j/\partial\phi < 0$; (iii) it rises in the uncontributed network cost γ , $\partial\Delta p_j/\partial\gamma > 0$. In a restructured state, where the incumbents' energy charge is a wholesale price that a persistent new load moves along the supply curve by $\eta \geq 0$ per megawatt-hour and generation capacity is in no rate base, the change in the incumbent price is*

$$\Delta p_j^R = L\eta + \gamma(1 - \phi)(-\lambda\theta + 1)\frac{\lambda}{1 - \lambda} - k\lambda\theta \quad (8)$$

with k and γ the wires values, and the energy channel is nonnegative: when the vertically integrated utility has slack ($m < c$), the restructured incumbent loses the energy-channel dividend that the regulated incumbent receives.

Part (i) is the discounted special contract: a data center that negotiates a tariff below the class average has $\theta < 1$ and shifts cost toward the incumbents; a large-load tariff with minimum demand payments has $\theta > 1$ and shifts it away. Part (ii) is the contribution in aid of construction that a large-load tariff typically requires. Part (iii) is the new transmission and distribution that a load of several hundred megawatts can require. The restructured case says that the sign of the energy channel depends on the regime: a vertically integrated utility averages a cheap marginal megawatt-hour into everyone's rate, while a restructured utility passes through a wholesale price that a new load can only raise. Two things follow for the data. First, the wholesale term ηL is a change in a market-wide price that reaches every default-service customer in the market, treated and control counties alike, so a within-market comparison nets it out; what the regime split in Section 9 can test is the wires-only fixed-cost channel, under which the restructured incumbent should see the smaller dividend. Second, the case is not only about restructured states. A utility that purchases its energy under wholesale or contract prices rather than generating it, which is how most municipal and cooperative systems are supplied, has the restructured energy channel whatever its state's regime: there is no embedded generation cost for a cheap marginal megawatt-hour to average down, so a new load brings such a system only the wires dividend, and any premium it carries on the energy side is passed through. The ownership split in Section 9 is the test of that case.

4.5 *Measurement: the class that bills the new load*

The data report each class's average revenue per megawatt-hour, and the new load is itself billed in one of the classes. If it is billed in class I at a tariff p_D below the incumbent price of that class, the measured change in the class average mixes the pass-through with the new customer's own tariff. With $\mu = L/(q_I + L)$ the new load's share of the class's sales and $\delta = p'_I - p_D > 0$ its discount relative to the incumbents,

$$\Delta \tilde{p}_I = \Delta p_I - \delta \mu \quad (9)$$

Proposition 3 (Composition). *The measured change in the billing class's average price is below the pass-through to its incumbents by $\mu\delta$, a wedge that is negative and grows with the new load's share of the class, $\partial(\Delta \tilde{p}_I - \Delta p_I)/\partial \mu = -\delta < 0$.*

Data centers are billed as commercial customers by most utilities in the EIA's classification, which puts nonmanufacturing establishments in the commercial class, and as

industrial customers by some. The residential class bills no data center, so the residential price is the clean measure of the pass-through, and the residential estimate is the headline. A measured industrial or commercial change below the residential one is the composition wedge, not evidence that households were charged more than businesses.

4.6 Dynamics: regulatory lag

Rates change at rate cases, not continuously. If a rate case arrives with a constant hazard h per year, the share of the long-run change that is in place τ years after the opening is

$$s_\tau = 1 - (1 - h)^{\tau+1} \quad (10)$$

and the event-study coefficient at event time τ is

$$\beta_\tau = \Delta p_j \left(1 - (1 - h)^{\tau+1} \right) \quad (11)$$

Proposition 4 (Regulatory lag). *The event-study coefficients β_τ move monotonically toward the long-run change Δp_j , with year-to-year increments that shrink geometrically,*

$$\frac{\beta_{\tau+1} - \beta_\tau}{\Delta p_j} = h(1 - h)^{\tau+1} \quad (12)$$

and the remainder $1 - s_\tau = (1 - h)^{\tau+1}$ vanishes as τ grows.

The prediction is a profile that rises or falls toward a limit, not a step at the opening. A profile that is flat at zero means either that the long-run change is itself zero or that the lag is long relative to the window; a profile that moves in one direction and then returns is outside the model.

4.7 What the signs rest on, and the mapping to the data

Proposition 1's sign rests on three assumptions: (i) fixed costs are allocated by an embedded-cost study, so that the new load's payment toward them is proportional to its sales; (ii) the new load pays the class tariff, so that $\theta = 1$; and (iii) the incumbents' own sales do not change. Relaxing (ii) toward a discounted contract reverses the dividend when the discount is large enough that θk falls below $m - c + (1 - \phi)\gamma$; a large uncontributed network cost does the same; a larger up-front contribution attenuates any cost shift. Relaxing (iii) toward incumbents who reduce consumption in response to a higher price adds a second-order term that Appendix A shows carries the sign of the first-order change and does not reverse it.

The data objects map onto the model's as follows. The county's Δp_j is the pass-through at its serving utility, weighted by that utility's share of the county's sales (Section 5). The event is the county's first data center; λ_c is the first facility's load relative to its serving utility's sales, which is not observed directly and is calibrated below. The estimating equation is equation (5) in logs, with the regulatory-lag profile applied:

$$\log p_{cjt} = \alpha_{cj} + \psi_{jt} + \sum_{\tau \neq -1} \beta_{j\tau} \mathbf{1}[t - G_c = \tau] + \varepsilon_{cjt}, \quad \beta_{j\tau} = s_\tau \frac{\lambda_c \pi_c}{p_j}, \quad (13)$$

where G_c is the year of county c 's first data center, α_{cj} absorbs permanent differences in a county's price level, ψ_{jt} absorbs national class-by-year shocks, and $\beta_{j\tau}$ is the model's proportional price change at event time τ . Proposition 1 signs β by the sign of the premium, Proposition 3 says the billing class's β is below the residential one, and Proposition 4 says β_τ moves monotonically with τ . Section 6 returns to equation (13) with the data objects substituted.

4.8 Calibration

Table 1 gives the model magnitudes. The new load is a 100 MW facility at an 80 percent load factor, 700,800 megawatt-hours a year; a size the calibration takes as representative of a large hall of the 2010s; Appendix D varies it through the load-share distribution. Its share of sales, λ , is computed against the state-wide sales of the largest utility serving each treated county in the year before its first opening: the median is 3.5 percent, the 10th percentile 0.9 percent, and the 90th percentile 14.4 percent. The embedded fixed-cost share of price, $\kappa = k/p$, is 50 percent: Watten et al. (2026) put marginal costs at approximately half of average costs, building on the national marginal-cost and price figures in Borenstein and Bushnell (2022), whose own figures put private marginal cost nearer a third of the average residential price, so the share is if anything conservative. The uncontributed network cost of the new load as a share of price, $g = (1 - \phi)\gamma/p$, is 10 percent in the central case; the paper's own range for it runs from zero, for a load that contributes its full interconnection cost under the large-load tariffs E3 (2026) catalog, to a few tens of percent for a load that contributes nothing, informed by the distribution cost-causation chapters of Lazar et al. (2020). The rate-case hazard h is 33 percent per year, a case every three years; Joskow and Schmalensee (2024) record that a third of utilities went without a formal review over 1958 to 1972, a hazard several times lower, so Appendix D also shows the profile at lower hazards. The hazard applies to the fixed-cost terms; the energy term passes through fuel and purchased-power clauses between cases, and Appendix D reports the window share under

Table 1. Calibration of the model and its predicted price changes

Parameter or quantity	Value	Source
L : annual load of a 100 MW facility at 80 percent load factor (MWh)	700,800	Shehabi et al. (2024): EIA's 80 percent assumption for mining loads
Q : serving utility's sales, median treated county (MWh)	19,571,938	EIA-861, this paper
$\lambda = L/(Q+L)$: median (10th–90th percentile)	0.035 (0.009–0.144)	this paper
$\kappa = k/p$: embedded fixed-cost share of price	0.50	Watten et al. (2026); Borenstein and Bushnell (2022)
g : uncontributed network cost of the new load, share of price	0.10	E3 (2026); Lazar et al. (2020)
h : rate-case hazard per year	0.33	Joskow and Schmalensee (2024)
θ_C, θ_I : allocation intensity of a load billed on the commercial, industrial tariff	1.01, 0.47	class prices, Table 2, at $\kappa = 0.5$
$\bar{s}$: share of the long-run change in place, mean over years 0–5	0.69	model, equation (10)
Predicted $\Delta \log p$, log points, long run (years 0–5 window): slack system ($a = -0.5$), sales allocation	-3.1 (-2.2)	
marginal = embedded energy cost ($a = 0$), sales allocation	-1.4 (-1.0)	
tight system ($a = +0.5$), sales allocation	0.3 (0.2)	
tight system, discounted tariff ($\theta = 0.5$)	1.2 (0.8)	
tight system, large-load tariff ($\theta = 1.5$)	-0.5 (-0.4)	
marginal = embedded, load billed on the industrial tariff (θ_I)	-0.5 (-0.3)	
tight system, load billed on the industrial tariff (θ_I)	1.3 (0.9)	
Premium implied by the estimate at 100 MW, percent of price (point; 95 percent CI)	-17 (-76, 42)	inverting Proposition 1
under the Rambachan–Roth set, $\bar{M} = 0.5$	(-209, 190)	
at 10 MW (95 percent CI)	(-738, 407)	
at 30 MW	(-248, 136)	
at 300 MW	(-27, 15)	
Minimum detectable effect at the headline precision, log points (premium at 100 MW, percent)	2.0 (84)	$2.8 \times \text{s.e.}$

Notes: The predicted changes are for an incumbent class at the median load share λ over the treated counties, with $\kappa = k/p$ the embedded fixed-cost share of price, g the uncontributed network cost of the new load as a share of price, $a = (m - c)/p$ the marginal-minus-embedded energy cost premium as a share of price, and θ the allocation intensity. The implied premium inverts Proposition 1 at the median λ and the mean share of the long-run change in place over years 0 to 5. Source: EIA-861 and the sources cited in the table; authors' calculations.

that two-channel profile as well. The hazard is also, in practice, asymmetric: a utility earning above its allowed return does not file, and one earning below does, so a load that shifts cost onto incumbents reaches rates sooner than one that pays a dividend, and a six-year window is more informative against cost shifting than for a dividend. The paper has no filing data for its treated utilities and carries the asymmetry as a stated limitation.

The calibrated predictions span the two regimes, and Table 1 gives each as a long-run change and, in parentheses, as the years 0 to 5 average the event study reports, which is the long-run change times the mean share 0.69 of it in place over that window. With a sales allocator and a system with slack ($a = -0.5$, the marginal megawatt-hour at half the average, where $a = (m - c)/p$ is the energy-cost premium as a share of price), an incumbent class's price falls by 3.1 log points in the long run at the median load share; with the marginal energy cost equal to the embedded cost it falls by 1.4; with a tight system ($a = +0.5$) it rises by 0.3. A tight system combined with a discounted tariff ($\theta = 0.5$) raises the price by 1.2 log points; a tight system combined with a large-load tariff ($\theta = 1.5$) lowers it by 0.5; a load billed on the industrial tariff in a tight system raises it by 1.3. At the 90th percentile of the load share the magnitudes are four

times larger: the sales-allocation benchmark with marginal cost at embedded cost is -5.7 log points and the tight-with-discount case is 5.2 . Section 7 holds the window-scaled values against the estimate, and Section 10 inverts Proposition 1 to report the premium the estimate implies at several assumed loads.

The model does not speak to where data centers locate, to market power in wholesale markets, to emissions, or to capacity additions beyond the network cost γ ; the literature on each is cited in Section 2 and left there.

5 Data

The paper rests on two new constructions: a county-level panel of retail electricity prices by customer class from 2001 to 2024, and a list of data-center facilities with opening dates drawn from a federal register. This section describes both and the sample they define. Appendix B gives the arithmetic, the crosswalks, and the counts step by step.

5.1 *County retail prices from Form EIA-861*

Form EIA-861 reports, for every electric utility and every state it serves, annual revenue (thousands of dollars), sales (megawatt-hours), and customers for the residential, commercial, industrial, and transportation classes, with revenue and sales broken out by the service the utility provides: bundled (Part A), energy-only (Part B), delivery-only (Part C), and bundled service by an adjustment entity (Part D). The Administration’s own state average price is the ratio of revenue to sales over the parts that do not double count a kilowatt-hour, and the utility-by-state price used here follows that arithmetic exactly. The files are public domain and were retrieved for every year from 1990 to 2024; the construction reproduces the county series the authors’ data library already carries for 2012 to 2024, built from the same files by the same arithmetic, to within floating-point error on every county-year-class cell and extends it back to 2001, the first year the form collected a county list.

The county is a list of service territories. Each utility names the counties it serves in a state, without allocating its sales among them, so the county price is the sales-weighted average over the utilities that list the county of each utility’s state-wide class price, weighted by the utility’s state-wide class sales. A county served by one utility carries that utility’s state average exactly; a county served by several carries an average weighted by their state-wide sizes rather

than their load inside the county. The county price therefore moves only when a serving utility's state-wide average moves, and the estimand is the effect of a county's first data center on the rates of the utilities that serve it, which is the object a rate case sets and the object the model describes. The name-to-FIPS crosswalk matches the utilities' county strings to Census county codes, with the misspellings, the Virginia independent cities, and the retired codes handled by rule; Appendix B reports its coverage and residue.

Three exclusions follow from the form. Texas is dropped: in the ERCOT market the competitive retailers that make most sales file no service territory, so the county construction covers about a third of the state's sales after 2002 and a Texas county price describes only its cooperatives and municipal utilities. The District of Columbia is dropped as a single-county unit. Utilities that file the short form from 2012 on report totals only, so they enter the all-classes price and are excluded from the three class prices; before 2012 every utility filed the long form, so the class prices are complete in the early years and miss the small utilities in the later ones. Alaska's rural utilities, with prices several times the national level, stay in the sample; in logs they are ordinary observations, and in the levels specification they carry most of the variance.

Table 2 reports the sample. The residential price in treated counties in the years before their first opening averaged 10.50 cents per kilowatt-hour against 11.04 in counties that never receive a facility; the commercial and industrial prices show smaller gaps in the same direction. Treated counties are served by cheaper utilities than untreated ones before anything opens, which is what sitting on the price of power would produce and is the selection the design must handle. The residential customer in a treated county used 11,732 kilowatt-hours a year before the opening, on a bill of \$1,280 (total residential revenue over total customers; the product of the two means is lower because each is a mean over county-years), the base for the dollar translation in Section 7.

5.2 A facility list with dates from Clean Air Act records

A data center of any size runs diesel generators for backup, and a stationary diesel generator above a small threshold needs an air permit from the state, which reports the facility to the Environmental Protection Agency's Integrated Compliance Information System for Air (ICIS-Air). The facility list is built from the public ICIS-Air download of 280,208 facilities nationally, with the Facility Registry Service for county codes and industry codes across

Table 2. Summary statistics: county retail prices by customer class, 2001–2024

Class	Treated, pre-opening	Never treated	S.d.	Treated counties	Never-treated counties	Counties	County-years
Residential	10.50	11.04	3.98	128	2,445	2,573	61,222
Commercial	8.69	9.26	3.58	128	2,445	2,573	61,294
Industrial	6.33	6.53	2.66	128	2,437	2,565	60,754
All classes	8.63	9.08	3.79	128	2,445	2,573	61,470

Notes: Prices are the county’s sales-weighted average over the utilities serving it of each utility’s state-wide class average revenue per kilowatt-hour, in cents. Treated counties are those with a dated first data center in the estimation sample (always-treated counties dropped); never-treated counties have no facility in the EPA list or the Global Data Center Map. Texas and the District of Columbia are excluded; county-years count every county-year with a price, and the estimation samples in Table 4 are a few rows smaller. Source: Form EIA-861; authors’ calculations.

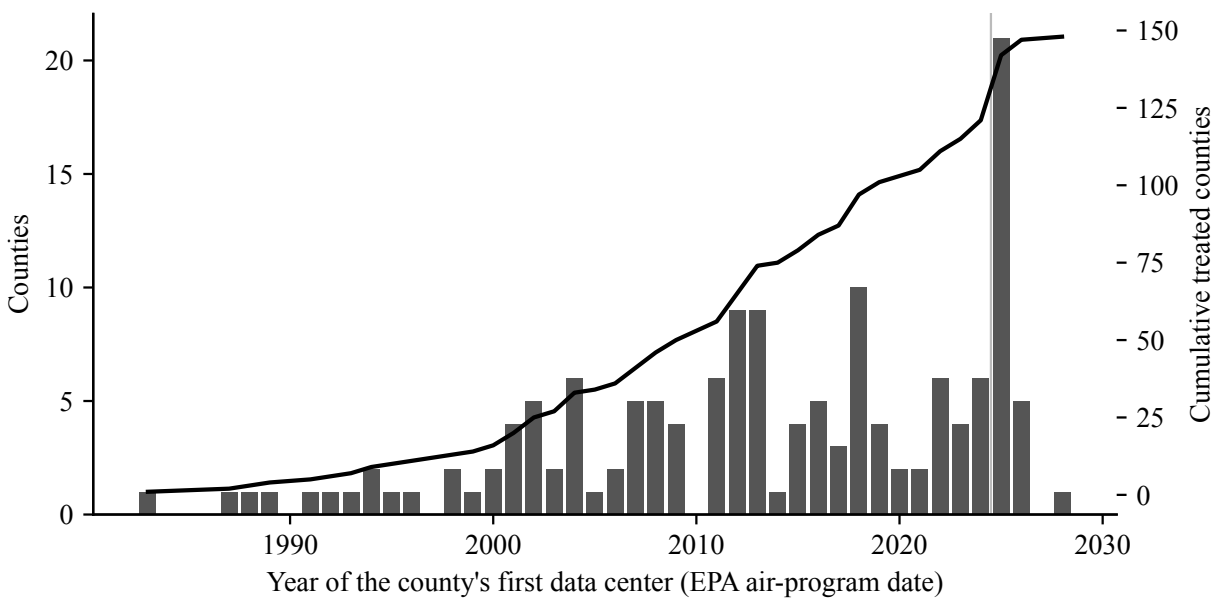
programs. A facility enters the list in one of three tiers: it carries NAICS 518210 (data processing, hosting, and related services) in ICIS-Air or in the registry’s cross-program industry list; its name says data center or colocation, or carries a cloud provider’s site code beside that provider’s name; or its name is that of a pure-play colocation operator or a hyperscale cloud operator. Telecommunications switching sites are excluded unless their name says data center; 7 cryptocurrency miners coded 518210 are kept and flagged. The rule was written before the known-cases test below and was not tuned to it. The list holds 623 facilities in 182 counties and 44 states; Appendix B lists the counts by tier, operator, and state.

The opening proxy is the facility’s air-program history. Each air program at a facility carries a begin date, and the first compliance evaluation carries a date, back to the 1970s. The complication is that ICIS-Air replaced its predecessor system in October 2014, and every program record migrated then carries a load stamp (October 19, 2014 for 213,627 program rows nationally, with a handful of later state reloads) rather than its true start. The dating rule treats a stamp as proof that the facility existed by that date and nothing more: a facility is dated by the earliest record that beats its earliest stamp, either an informative program begin date (date quality “permit”) or its first compliance evaluation (“evaluation”); a facility with only stamped records is censored and carries an upper bound. Of the 179 counties with a facility that appear in the price data (the rest of the 182 are in Texas or the District of Columbia), 62 are dated by a permit and 90 by an evaluation, 152 in all; 24 carry only a stamp (15 censored at the 2014 migration, 9 at a later state reload) and are excluded from the treated group, and 3 carry no record; a begin date after 2024 belongs to a facility permitted ahead of construction, which is why the latest cohorts run to 2028. Of the dated counties, 148 have the class prices the estimation needs (4 lack them in every year), and those are the treated counties. Table 3 and Figure 1 show when the first openings fall.

Table 3. Counties by the year of their first data center and by date quality

Year of the county's first data center	Counties
2001 or earlier (always treated in the price panel; dropped)	20
2002–2011	36
2012–2018	41
2019–2024	24
2025 or later (not yet treated in the price panel)	27
Counties with a dated first opening	148
dated by an air-program begin date	62
dated by a first compliance evaluation	90
Counties whose facilities carry only the 2014 migration stamp (excluded)	15
Counties whose facilities carry only a later bulk stamp (excluded)	9
Counties whose facilities carry no date (excluded)	3

Notes: A county's first data center is the earliest dated facility in the EPA list. Counties whose facilities carry only a bulk-load date stamp, or no date, are excluded from the treated group and from the clean controls. The full year-by-year list is Table B2. Source: EPA ICIS-Air and Facility Registry Service; authors' calculations.

Figure 1. First openings by year, and the cumulative count of treated counties

Notes: Bars are counties by the year of their first dated facility; the line is the cumulative count. The vertical line marks 2024, the last year of the price panel: counties dated after it contribute pre-opening years only. Source: EPA ICIS-Air; authors' calculations.

The proxy was tested against 31 documented hyperscale openings whose county and year are public: 15 were found under the operator’s name in the right county, and the EPA date was inside the documented in-service year for 4, one to three years late for 7, more than three years late for 2, and earlier than the in-service year for 1, where an air permit was issued before the hall went into service; the mean gap was 1.2 years (Table B1). Of the 16 misses, 9 counties hold a facility under a name the operator rule cannot read, a shell company or a contractor, so the county is treated correctly and only the label is lost; 7 counties hold no facility at all, three of them in Oregon and one each in Ohio, Utah, Illinois, and New York, which says those states keep minor-source generator permits out of the federal system; all seven are in the Global Data Center Map and so are excluded from the clean control group. Iowa and North Carolina have both hits and misses, so part of the incompleteness is the list and not the state. The direction of each error is known: a late date puts the first years of a true effect into the pre-period and flattens the event study; an early date, from a permit issued ahead of service, puts untreated years into the post-period and attenuates it; a missed facility makes a treated county look untreated and attenuates the estimate toward zero.

The County Business Patterns establishment counts for the industry (NAICS 518210 and its predecessors, 1986 to 2023) are a second, agency-file dating that the battery uses to re-date openings by the first year a county shows an establishment with 50 or more employees. It is a loose check: hosting is only part of the industry, and most large metropolitan counties show such an establishment in 1986, the first year of the files, so re-dating turns them into always-treated counties and the check runs on the remainder. The Global Data Center Map, a public compilation of 8,818 U.S. facilities with addresses and no dates, is used as a second universe to define the control group: a never-treated county must have no facility in the EPA list and none in the map.

5.3 The estimation sample

The panel has 2,573 counties in 48 states over 2001 to 2024 in the residential class. Of the 148 counties with a dated first opening, 20 were first dated in 2001 or earlier and are dropped from every estimate, as units with no pre-period; Loudoun County, Virginia, the paper’s opening example, is among them, as are the other early clusters, so the estimates describe counties that received their first facility after the price panel begins; 27 are first dated after 2024 and stay as not-yet-treated units that contribute pre-opening years only; 128 treated counties remain, of

which 101 have at least one post-opening year. The clean control group is 2,445 counties with no facility in either list. The cohorts fall in three waves: 36 counties first dated 2002 to 2011, 41 in 2012 to 2018, and 24 in 2019 to 2024 (Table 3); the median first opening falls in the middle wave.

Two heterogeneity variables are fixed per county. The regulatory regime is the state's, from the lists above: 39 of the 128 treated counties in the estimation sample are in restructured states. The ownership of the county's largest serving utility is the modal class over the panel, investor-owned for 116 treated counties, municipal or public for 8, and cooperative for 4. The load-share proxy that the dose-response in Section 9 uses is a 100 megawatt load against the largest serving utility's sales in the county's first panel year.

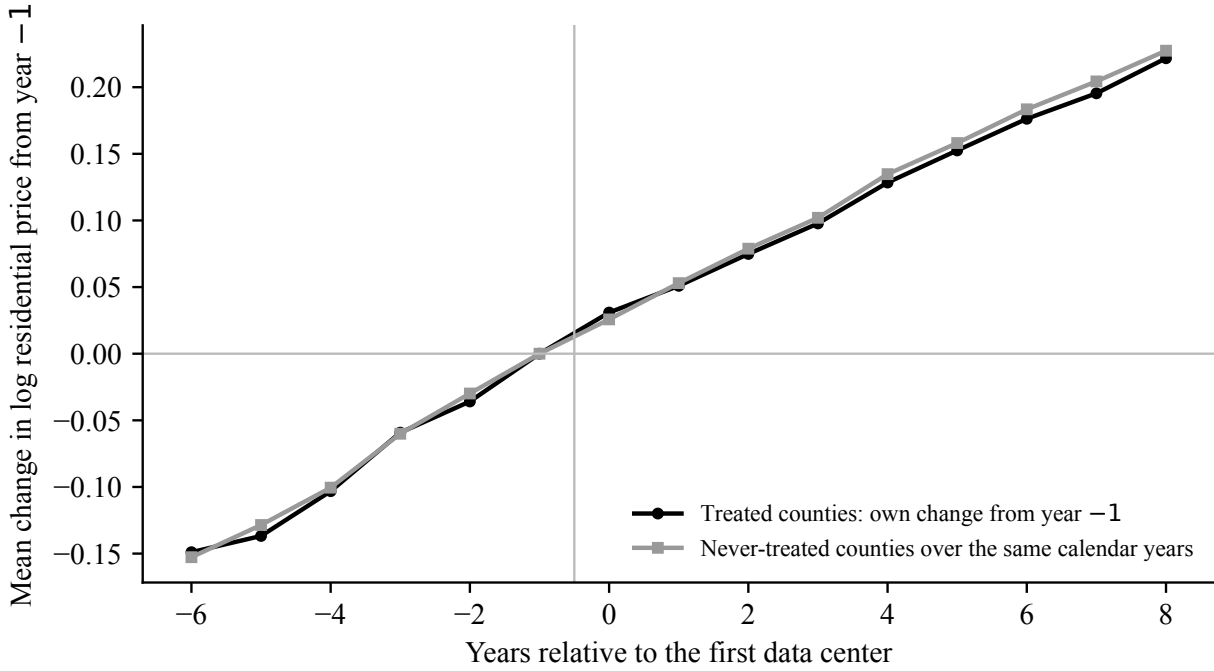
Figure 2 shows the raw residential price paths before any estimation: for each treated county and event year, its own change in log price from the year before the opening, and the never-treated counties' mean change over the same two calendar years, each averaged over treated counties by event year, so the gap between the lines is a simple difference-in-differences with no regression. The two paths move together before the opening and stay close after it. Sections 6 and 7 put standard errors on that picture.

6 Empirical strategy

6.1 *Estimand*

The parameter is the average effect, on counties that receive their first data center, of that opening on the log of the county's class average retail price in each of the years after it; the headline is the mean over the first six years (event years 0 to 5), the window pre-specified in the Identification Memo (the design document committed to the replication package before any estimate was run), and the mean over years 6 to 8 is reported beside it because the model's regulatory-lag profile keeps moving. It is an average treatment effect on the treated, in log points, for the counties whose first facility the EPA list dates between 2002 and 2024; it converts to cents per kilowatt-hour at the treated counties' pre-opening mean price and to dollars per household-year at their mean residential consumption. The parameter is the effect of a county's first data center on the rates of the utilities that serve it (Section 5), which is the object a rate case sets. The model's Proposition 1 says the parameter is $s_\tau \lambda_c \pi_c / p_j$, and the dose-response in Section 9 estimates its slope in λ .

Figure 2. Residential prices around the first data center: treated counties and never-treated counties on the same calendar years



Notes: Each treated county's change in log residential price from the year before its first opening, averaged by event year; the never-treated line is the mean change of the never-treated counties over the same pair of calendar years, averaged over the same treated counties, so the two lines share the same calendar composition and their gap is a raw difference-in-differences. Source: Form EIA-861 and EPA ICIS-Air; authors' calculations.

6.2 Design

The variation is the staggered year in which counties receive their first facility, compared with counties that have not yet received one and with counties that never do. The estimating equation is equation (13): the log price of class j in county c and year t on county-by-class fixed effects, which absorb permanent differences in a county's price level from fuel mix, ownership, and density; class-by-year fixed effects, which absorb national price shocks from fuel prices, interest rates, and the business cycle; and event-time indicators for every year relative to the first opening except the year before it, which is the reference. The window reported is six years before to eight years after.

Under staggered adoption the two-way fixed-effects estimator of equation (13) averages comparisons that use already-treated counties as controls with weights that can be negative (Goodman-Bacon, 2021; de Chaisemartin and D'Haultfœuille, 2020), so the headline is estimated three ways that do not. The Sun and Abraham (2021) interaction-weighted estimator estimates

each cohort's path against the never-treated counties and averages the cohorts by their share of the treated; it is the headline. The Callaway and Sant'Anna (2021) estimator computes each cohort-by-year effect against the counties not yet treated in that year and aggregates by event time; it uses a different comparison group and a different weighting. The stacked estimator of Cengiz et al. (2019) builds, for each cohort, a clean experiment of the cohort and the counties not treated anywhere in its window, stacks the experiments with experiment-specific fixed effects, and estimates one path. The two-way fixed-effects estimator is reported beside them as the benchmark, with the Goodman-Bacon (2021) decomposition of its weights in Appendix C. The post-window mean and its standard error come from the joint covariance of the event-time coefficients in each case.

Standard errors are clustered by state. Retail rates are regulated by state commissions, the EIA construction attributes a utility's state-wide price to every county it serves, so errors are mechanically correlated across the counties a utility serves within a state, and the regime and policy shocks that threaten the design are state-level. Because Texas is dropped and some specifications have few treated clusters, a wild-cluster bootstrap with Webb weights (Cameron et al., 2008) is reported for the static specification, with the caution that MacKinnon and Webb (2018) show the wild-cluster bootstrap itself under-rejects when the treated clusters are few, so on a cell with a handful of treated states a large bootstrap p-value is a statement that the cell cannot be tested, not that it is zero. The baseline rows are unweighted, so that the estimand is the effect on a county's residents' price; a version weighted by residential customers is in the battery, under the Callaway–Sant'Anna estimator, which takes weights.

6.3 Threats

The identifying assumption is that, absent the opening, treated counties' log prices would have moved like not-yet-treated and never-treated counties' within the same class and year, and that the opening was not anticipated in rates. The threats are listed in the order of the damage they would do.

Selection on trajectories is the first. Cheap and stable power is a site criterion, and Table 2 shows treated counties were served by cheaper utilities before any opening. If data centers select utilities whose rates were already on a different path, the plausible direction is toward utilities with low and falling rates, which would bias the estimate downward, toward finding a dividend

that is really selection. Anticipation and late dates are the second: a utility that builds generation for a load it has contracted raises rates before the facility opens, and the EPA date runs late in most of the known cases, so part of a true effect lands in the pre-period and the post-opening estimate is attenuated. Staggered bias in the two-way fixed-effects estimator is the third, and is why it is a benchmark and not the headline. Misclassified controls are the fourth: a facility the list misses, in a state that keeps its generator permits out of the federal system or under a name the rule cannot read, makes a treated county a control and attenuates the estimate toward zero.

Composition is the fifth threat, and is the model's Proposition 3: the class that bills the new load shows a measured change below the true pass-through, so the industrial and commercial estimates are read with that wedge in mind and the residential estimate is the clean one. State policy shocks that coincide with openings, such as a renewable portfolio standard, a fuel-cost pass-through, or a restructuring, are the sixth. The seventh is the unit: the county price is a utility average, so untreated counties served by a treated utility carry the treated utility's price and a within-utility comparison is zero by construction, which would attenuate a county design toward zero. The last is the whole-thesis objection: that the facility list is incomplete and late, so that an estimate near zero is attenuation rather than absence.

6.4 *Diagnostics*

Each threat has a test that could fail. Selection on trajectories: the five pre-opening coefficients and their joint test, with the Roth (2022) caution that passing is weak evidence, the Rambachan and Roth (2023) confidence sets under the relative-magnitudes restriction, in which the period-to-period change in the trend violation after the opening is at most a multiple of the largest such change before it, and a never-treated control group matched to the treated counties on their pre-period price level and growth. Anticipation and late dates: the coefficients at -2 and -1 against those at -6 to -3 ; the cohorts dated by a permit rather than an evaluation; re-dating from the County Business Patterns; a placebo that moves every opening three and five years earlier on pre-opening data alone; and, because the known cases say the EPA date is late in most cases and early in one, the design re-run with every opening shifted one and two years earlier.

Staggered bias: the three robust estimators against the benchmark, and the decomposition. Misclassified controls: the clean control group against all counties without an EPA facility; if the misses are many, including them as controls should move the estimate toward zero. Composition:

the residential-minus-industrial and residential-minus-total gaps, in the full sample and in every dose subset. State shocks: state-by-year fixed effects, which identify the effect from the timing of openings within a state. The unit: a control group restricted to counties whose largest utility serves no treated county; one treated county per utility; and the design re-run at the utility-by-state unit. The whole-thesis objection: a first stage, the same event study on the serving utilities' commercial and industrial sales, which says whether the first opening was a load the utility could see; the Clean Air Act class split, which tests the model's gradient on the better-measured facilities; and the known-cases test, which bounds the lateness. The diagnostics are reported with the results, so that the reader can judge the design before reading what it finds.

7 Results

7.1 *The headline estimates*

Table 4 reports the mean effect of a county's first data center on log class prices over the six years after the opening, by estimator and by class, for 128 treated counties, of which 27 dated after 2024 contribute pre-opening years only. In the residential class the Sun–Abraham estimate is -0.4 log points with a standard error of 0.7 , a 95 percent interval from -1.8 to 1.0 log points. The Callaway–Sant'Anna estimate, which compares each cohort with the counties not yet treated rather than with those never treated, is -0.3 (0.7); the stacked estimate is -0.7 (1.0); the two-way fixed-effects benchmark is -0.5 (1.1). The four agree within a standard error, and none is distinguishable from zero; the two-way fixed-effects figure here is the mean of its event-time coefficients over years 0 to 5, and the static coefficient is reported with the decomposition below. The agreement reflects the size of the never-treated group; the decomposition below says why.

Of the 128 treated counties, 101 have at least one post-opening year and contribute to the post-window mean; the 2019 to 2024 cohorts contribute at most five post-opening years each, so the later event times are estimated from the earlier cohorts. The never-treated group is 2,445 counties, and the not-yet-treated group that the Callaway–Sant'Anna estimator uses shrinks as cohorts are treated, which is why its standard errors at long horizons are wider than Sun–Abraham's. The all-classes and industrial samples differ from the residential one by a few counties whose class prices are missing in some years.

Figure 3 shows the residential path. The five pre-opening coefficients are small and of mixed sign, the largest at year -6 ; the joint test rejects them at $p = 0.05$ under Sun–Abraham,

Table 4. The effect of a county's first data center on log retail prices, years 0 to 5, by class and estimator

Estimator	Residential	Commercial	Industrial	All classes
Sun–Abraham	-0.41 (0.72)	-1.19 (1.22)	-0.42 (1.25)	-0.61 (0.94)
Callaway–Sant’Anna	-0.28 (0.75)	-1.06 (0.98)	-0.14 (1.19)	-0.56 (0.84)
Stacked	-0.73 (1.01)	-1.08 (1.34)	-0.41 (1.48)	-0.69 (1.14)
Two-way FE	-0.46 (1.10)	-0.70 (1.45)	-0.21 (1.66)	-0.49 (1.29)
Pre-trend joint test, p (Sun–Abraham)	0.05	0.32	0.01	0.09
Pre-trend joint test, p (Callaway–Sant’Anna)	0.13	0.50	0.06	0.25
Pre-trend joint test, p (Stacked)	0.01	0.06	0.00	0.01
Pre-trend joint test, p (Two-way FE)	0.31	0.20	0.09	0.11
Years 6 to 8 (Sun–Abraham)	-0.97 (1.51)	-2.03 (2.25)	0.56 (2.36)	-0.84 (1.78)
Treated counties	128	128	128	128
Never-treated counties	2,445	2,445	2,437	2,445
County-years	61,222	61,294	60,750	61,470

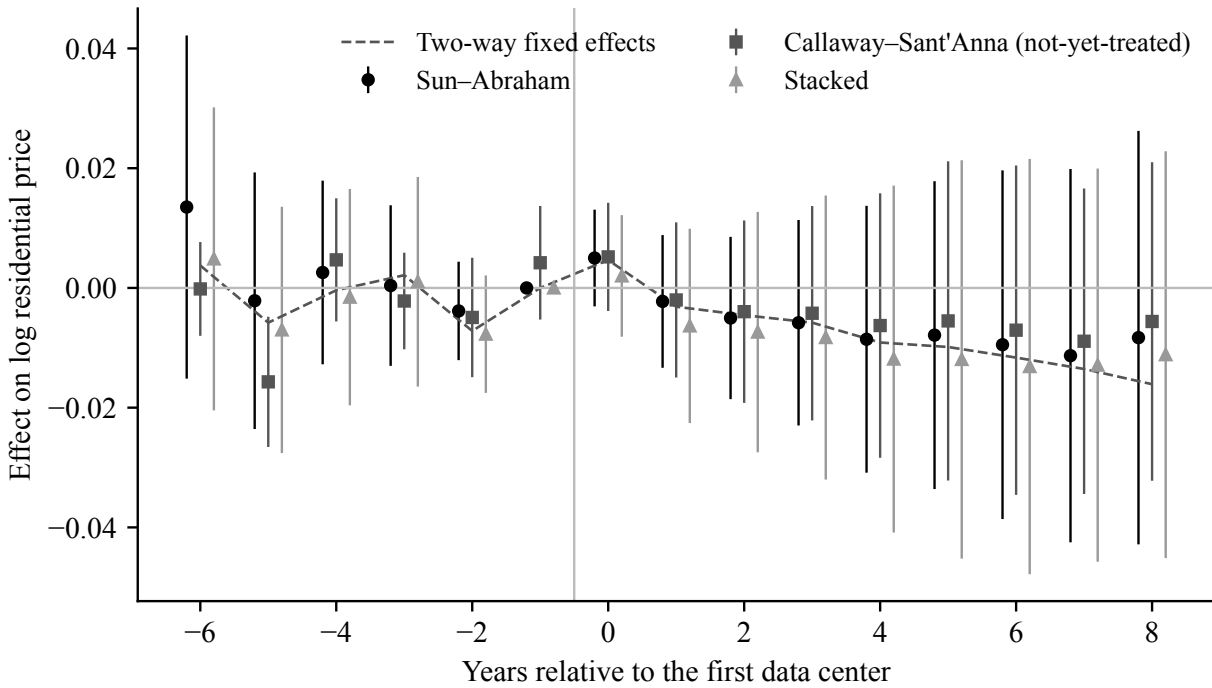
Notes: Each cell is the mean of the event-time coefficients for years 0 to 5 after the first opening, in log points, from an event study of the log county price on county and year fixed effects with a window of six years before to eight years after; standard errors clustered by state in parentheses. Sun–Abraham uses never-treated counties as the comparison; Callaway–Sant’Anna uses not-yet-treated counties; the stacked estimator uses clean controls within each cohort’s window; two-way FE is the benchmark. The pre-trend test is a joint Wald test that the coefficients for years -6 to -2 are zero. Source: Form EIA-861 and EPA ICIS-Air; authors’ calculations.

which is borderline, and at $p = 0.007$ under the stacked estimator, which is a rejection. The stacked leads are noisier because each stack has fewer controls; Section 8 reports the confidence sets that allow violations of parallel trends of the size the leads show. After the opening the coefficients drift down, from 0.5 log points in the opening year to -0.6 in year 3, -0.8 in year 5 and -0.8 (1.8) in year 8, without leaving the band at any horizon. The mean over years 6 to 8 is -1.0 (1.5).

7.2 The first stage: what the serving utilities’ sales show

A price can only move if the load moved the utility, so the same event study is run on the log of the serving utilities’ commercial and industrial sales, the county row’s own sales field,

Figure 3. Event study: the effect of a county's first data center on the log residential price



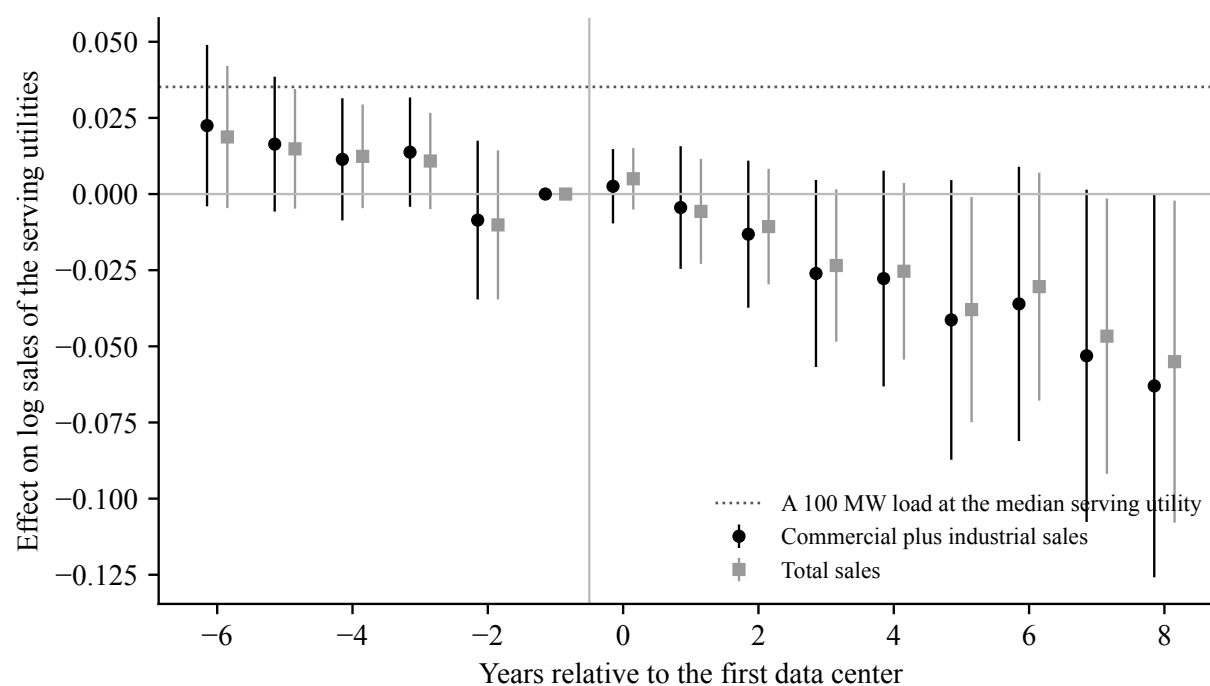
Notes: Coefficients on years relative to the first opening (year -1 is the reference) with 95 percent confidence bands, standard errors clustered by state, for the Sun-Abraham, Callaway-Sant'Anna, and stacked estimators; the dashed line is the two-way fixed-effects path. Source: Form EIA-861 and EPA ICIS-Air; authors' calculations.

which is the model's Q . Over years 0 to 5 the first opening is associated with a change of -1.8 log points (1.3) in those sales and -1.6 (1.1) in total sales, point estimates with the wrong sign for a new load. The 95 percent interval on the first, -4.4 to 0.8 log points, excludes the 3.5 percent of sales a 100 MW facility would add at the median utility and cannot exclude the few tenths of a percent a facility of a few megawatts would add. The pre-opening test gives $p = 0.17$. At the median serving utility that upper bound is a load of about 21 MW, so the typical first facility was well below the calibration's 100 MW; it is not a load the design can see in the utility's sales, and the price null that follows is a null on a small treatment. The residential class's own quantities, which can move the residential average revenue with no change in rates, move at the margin and in opposite directions: residential sales per customer change by 2.5 log points (1.3) and residential customer counts by -3.2 (1.6), each about two standard errors from zero, leaving total residential sales about unchanged. A rise in consumption per customer lowers the average revenue per kilowatt-hour where fixed charges and declining blocks are in the tariff, so part of a residential price change in these data can be composition within the class rather than a

Table 5. The first stage: the serving utilities' sales around a county's first data center

Outcome (log of the serving utilities' sales)	Years 0 to 5 (log points)	Years 6 to 8	Pre-trend p	Treated	Never treated
Commercial plus industrial sales	-1.84 (1.32)	-5.07 (2.72)	0.17	128	2,437
Total sales	-1.63 (1.08)	-4.40 (2.27)	0.33	128	2,437
Residential sales per customer	2.55 (1.32)	—	0.05	128	—
Residential customers	-3.20 (1.59)	—	0.34	128	—
Treated counties whose first facility is the serving utility's first	64 of 128				
Distinct serving utilities among treated counties	63				

Notes: Sun–Abraham event studies of the log of the serving utilities' state-wide sales (the county row's sales field) on county and year fixed effects, the same design as Table 4; standard errors clustered by state. A county's event is its serving utility's first when no other county for which that utility is the largest utility was dated earlier. Source: Form EIA-861 and EPA ICIS-Air; authors' calculations.

Figure 4. Event study: the serving utilities' commercial-plus-industrial sales

Notes: Coefficients on years relative to the first opening (year -1 is the reference) with 95 percent confidence bands, standard errors clustered by state; the dotted line is the sales a 100 MW load would add at the median serving utility. Source: Form EIA-861 and EPA ICIS-Air; authors' calculations.

rate change; the magnitudes here are small, and the caveat is stated rather than resolved. The count of events the model treats as new load is also smaller than the count of counties: 64 of the 128 treated counties (50 percent) are the first dated facility among the counties their largest utility serves, and the 63 distinct serving utilities are the number of price paths the design actually compares. The utility-unit design in Section 8 is the version that counts each path once. Table 5 and Figure 4 report the first stage.

Table 6. The residential estimate in cents per kilowatt-hour and dollars per household-year

	Value
Effect on log residential price, years 0 to 5 (log points)	-0.41
Pre-opening residential price in treated counties (cents per kWh)	10.50
Effect in cents per kWh	-0.043
95 percent confidence interval, cents per kWh	[-0.191, 0.105]
Annual residential consumption per customer, treated counties (kWh)	11,732
Effect in dollars per household-year	-5.04
95 percent confidence interval, dollars per household-year	[-22.44, 12.36]
Upper bound, dollars per household-month	1.03

Notes: The log-point estimate is the Sun–Abraham mean over years 0 to 5 (Table 4); the pre-opening price and consumption are means over treated counties’ pre-opening years, consumption as total residential sales over total residential customers. Source: Form EIA-861 and EPA ICIS-Air; authors’ calculations.

7.3 *The magnitude, in cents and dollars*

Table 6 converts the residential estimate. At the treated counties’ pre-opening residential price of 10.50 cents per kilowatt-hour, the point estimate is -0.043 cents, with a 95 percent interval from -0.19 to 0.11 cents. At the treated counties’ mean residential consumption of 11,732 kilowatt-hours a year, that is a saving of \$5.04 on an annual bill, with an interval from a saving of \$22.44 to an increase of \$12.36. Under parallel trends, then, the interval excludes an increase of more than \$1.03 a month on a bill of about \$1,280 a year. Under the Rambachan and Roth (2023) restriction that the post-opening violation of parallel trends is at most half the largest pre-opening one, the set widens to -5.0 to 4.5 log points, which at the same price and consumption is about \$4.67 a month at the top; Section 8 reports the full set of restrictions. The paper’s bound is the one under the weaker assumption.

Two published magnitudes are the benchmarks. Garimella (2026) reports, for ZIP codes in 24 states over 2014 to 2024, “a \$2.09 per month increase in residential electricity bills” at the ZIP level, marginally significant, and at the utility-territory level “a household-weighted estimate of $-\$2.00$ ($p = 0.605$)”; this paper’s interval under parallel trends, from a saving of \$1.87 to an increase of \$1.03 a month, excludes the first and includes the second, and the set under modest violations includes both. Watten et al. (2026) estimate at the state level that “for every 10% increase in data center capacity average residential retail prices fell by approximately 0.4% on

average”; a county’s first opening is a different treatment from a 10 percent increase in capacity, and a decline of four tenths of a percent sits inside the county interval, so the two results are consistent with each other and neither rejects the other.

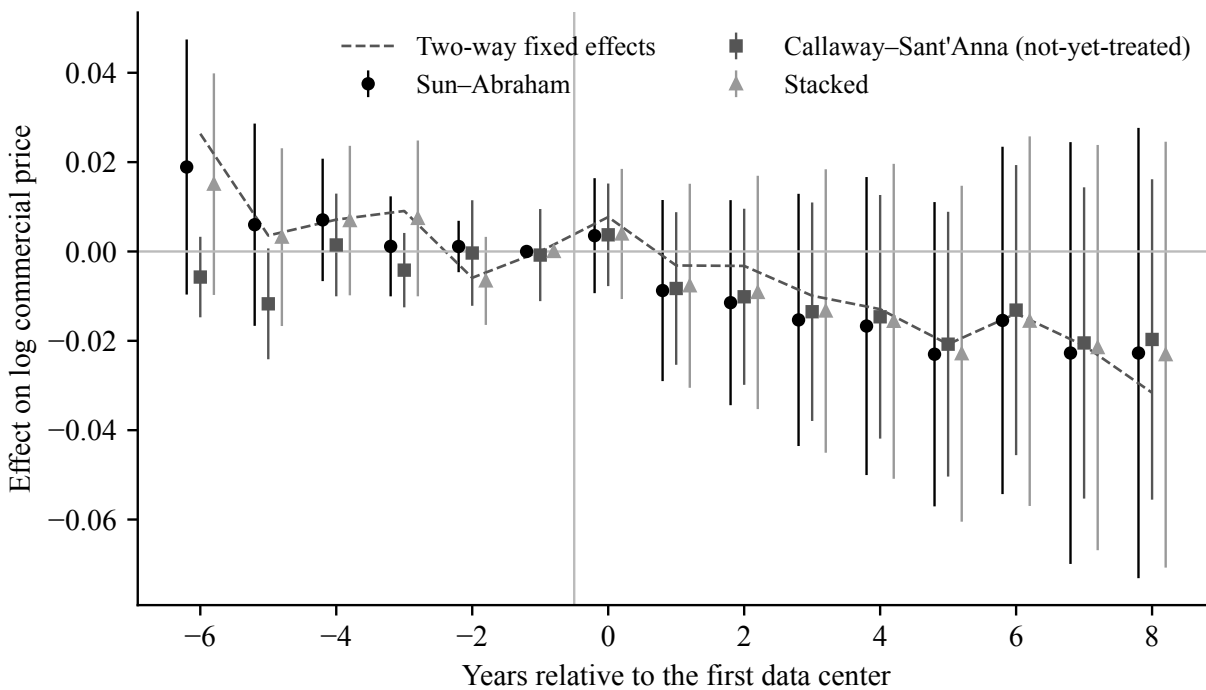
7.4 What the bound says against the model

The Identification Memo pre-specified the test: a null is informative if the interval excludes the calibrated magnitude for the cost-shifting case, a sales allocator with no up-front contribution; the memo’s dated amendment, recorded before any estimate was run, fixed that case’s marginal energy cost at half again above the embedded one and the network cost at the central calibration. The comparison is made on the years 0 to 5 window, so each calibrated long-run change is scaled by the 0.69 of it the regulatory-lag profile puts in place over that window: the cost-shifting case is then 0.2 log points at the median load share, the neutral benchmark -1.0 , the slack dividend -2.2 , and the discounted-tariff case in a tight system 0.8 . Under parallel trends the interval, -1.8 to 1.0 , excludes only the slack dividend; under the Rambachan–Roth set it excludes none of them. The test the memo set was not passable at the precision the design has: the minimum detectable effect at the headline standard error is 2.0 log points, and at a 100 MW load and the median load share that is a premium of 84 percent of the retail price. Inverting Proposition 1 at that load, the estimate implies a premium of -17 percent of the retail price, with a 95 percent interval from -76 to 42 percent under parallel trends; Section 10 repeats the inversion at other loads and under the weaker trend assumption. The design bounds the premium only on an assumption about the load, and at the loads the first stage suggests it does not bound it at all. The splits in Section 9 are where the sign is examined, on the facility’s permit class, which the public record proxies coarsely, and on the ownership of the serving utility, which it measures well.

7.5 The other classes

The commercial class is the most negative: -1.2 log points (1.2) under Sun–Abraham and -1.1 (1.0) under Callaway–Sant’Anna (Figure 5). Most utilities bill a data center as a commercial customer, so Proposition 3 predicts that the measured commercial average falls by the new customer’s own discount times its share of the class, whatever happens to the incumbents’ rates. The wedge is a statement about the tariff the new load pays relative to the

Figure 5. Event study: the log commercial price

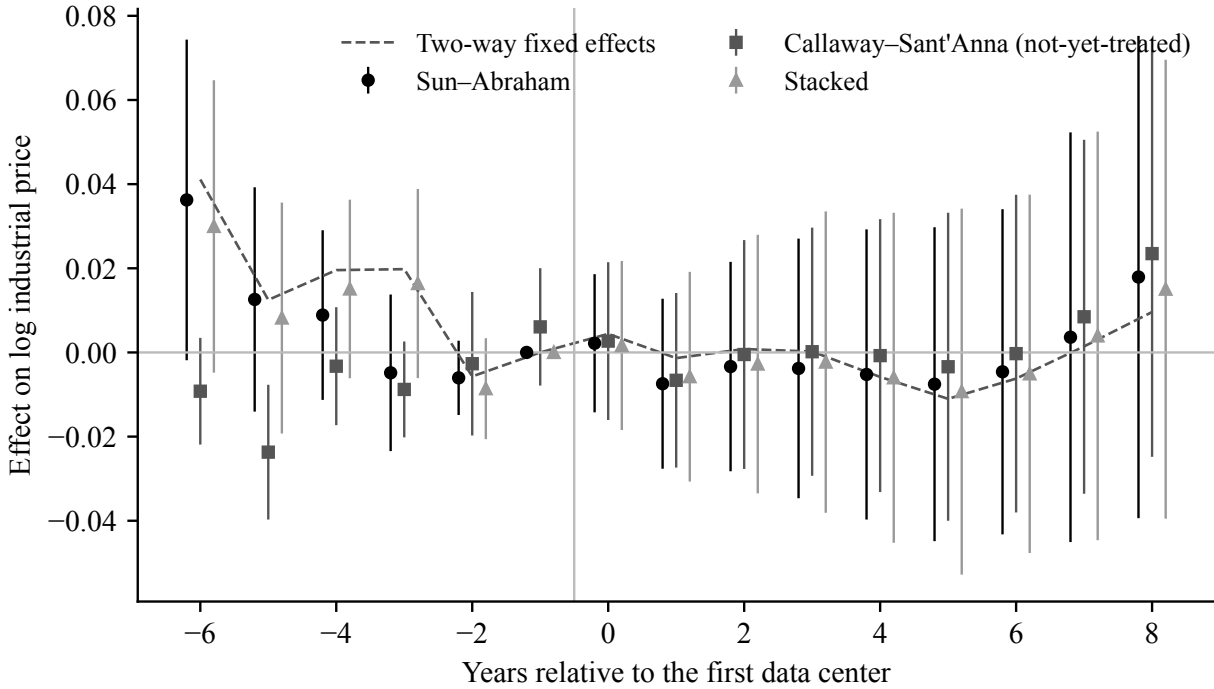


Notes: As Figure 3, for the commercial class. Source: Form EIA-861 and EPA ICIS-Air; authors' calculations.

class average, not about cost causation: a load-factor-justified lower rate and a discount that shifts cost look the same in these data. For the commercial estimate to be composition alone, the new load's discount would have to be 14 percent of the class price at a 100 MW load that is 8.7 percent of the class's sales, which is large, and the gap between the commercial and residential estimates is not precise enough to measure the wedge in any case. The industrial estimate is -0.4 (1.3) under Sun-Abraham and -0.1 (1.2) under Callaway-Sant'Anna (Figure 6); its pre-opening coefficients are rejected at $p = 0.006$, driven by a positive coefficient at year -6 , so the industrial path is read with less confidence than the residential one. The all-classes price, which includes the short-form utilities the class prices omit, moves by -0.6 (0.9).

The residential-minus-industrial gap, the model's composition statistic for a load billed as industrial, is 0.0 log points with a standard error of 1.4: at the first opening, no composition wedge is detectable, which is what a first facility of a few megawatts in a class of millions of megawatt-hours should produce. Under the general allocator of Appendix A the gap also carries a term of the opposite sign, because the residential class's embedded fixed cost per kilowatt-hour is higher than the industrial class's (k_j differs by class), so a pure dividend lowers the residential price by more; the gap is the net of the two and is read as such in Section 9.

Figure 6. Event study: the log industrial price



Notes: As Figure 3, for the industrial class. Source: Form EIA-861 and EPA ICIS-Air; authors' calculations.

7.6 Dynamics

Proposition 4 predicts a path that moves monotonically toward the long-run change, with increments that shrink. With a negative premium the prediction is a path that falls, and the residential path does fall, from the opening year to year 8, with no increment outside the band. If the profile is believed, the six-year average understates the long-run change: the years 6 to 8 mean, -1.0 log points, is about twice the years 0 to 5 mean. The calibration's rate-case hazard puts 0.69 of the long-run change in place over the headline window, which is the same arithmetic. At the calibrated hazard the share in place reaches 0.97 by year 8, so a long-run change of about one log point would show up as the tail does. The paper keeps years 0 to 5 as the headline because that window was pre-specified, and reports the tail beside it; the tail is estimated from the earlier cohorts alone and its standard error is correspondingly larger.

7.7 The benchmark and its weights

The static two-way fixed-effects coefficient for the residential class, one treated-by-post indicator on the full panel, is -0.9 log points (1.8); Table 4's two-way figure is the event-study

mean over years 0 to 5, and Table C1's is the static coefficient on the balanced subsample the decomposition requires. The Goodman-Bacon (2021) decomposition in Table C1 puts 0.981 of its weight on comparisons of treated counties with never-treated ones and 0.009 on the comparisons that use already-treated counties as controls, so the benchmark is close to the robust estimators here because the never-treated group is large relative to the treated one, not because staggering does not matter. The wild-cluster bootstrap on the static two-way specification, with 48 state clusters and Webb weights, gives $p = 0.64$ on 38 treated states, so the clustered inference is not understating the uncertainty at the headline.

7.8 *Date quality*

Among the cohorts dated by an informative air-program begin date rather than a first inspection, 61 counties (one permit-dated county is outside the estimation sample), the residential estimate is -0.8 (0.9), the all-classes estimate -2.7 (1.1), and the industrial estimate -3.3 (1.4). The better-dated subset shows a residential price that did not move and non-residential prices that fell. The spread has Proposition 3's sign, but at a first facility of a few megawatts the wedge the proposition produces is small: for the all-classes decline to be composition alone the new load would need a discount of 77 percent of the average price at a 100 MW load, and the industrial decline 28 percent of the industrial price. Those are not plausible tariffs, so the non-residential declines in this subset are larger than composition explains and are left as unexplained; they are not read as a dividend to incumbents either, since the residential price did not move. Re-dating the openings from the County Business Patterns, in Section 8, turns the residential point estimate positive on the counties it can re-date, which is a second reason to treat date quality as a limitation rather than a dose.

In sum, the first opening left the residential price within about five percent of where it would otherwise have been under modest violations of parallel trends and within about two percent under none; the serving utilities' sales show no load the design can see; and the design bounds the premium of the typical opening only on an assumption about its size.

8 Robustness

Table 7 reports the residential, commercial, all-classes, and industrial estimates under the specifications the Identification Memo pre-specified or added in its dated amendments (the

Table 7. The specification table: years 0 to 5, Sun–Abraham unless noted

Specification	Residential	Commercial	All classes	Industrial	Pre-trend p (res.)	Treated
Baseline (clean controls, state clusters)	-0.41 (0.72)	-1.19 (1.22)	-0.61 (0.94)	-0.42 (1.25)	0.05	128
All non-EPA counties as controls	-0.41 (0.71)	-1.10 (1.19)	-0.58 (0.91)	-0.34 (1.22)	0.05	128
Utility-clean controls	-0.13 (0.95)	-0.69 (1.50)	0.03 (1.16)	0.94 (1.47)	0.05	128
Permit-dated cohorts only	-0.82 (0.92)	-3.35 (1.11)	-2.67 (1.11)	-3.29 (1.41)	0.26	61
Cohorts re-dated from CBP employment ≥ 50	0.90 (0.79)	0.19 (1.75)	0.51 (0.91)	0.64 (1.33)	0.00	50
Cohorts 2012 and later	0.02 (0.85)	-0.61 (1.24)	-0.62 (1.01)	-0.56 (1.38)	0.12	92
Cohorts before 2012	-0.98 (1.17)	-1.99 (2.01)	-0.55 (1.59)	-0.09 (2.27)	0.31	36
One treated county per utility, utility-clean controls	0.19 (1.01)	-0.15 (1.62)	0.60 (1.31)	1.39 (1.71)	0.01	61
Dropping Virginia	-0.64 (0.65)	-1.47 (1.13)	-0.88 (0.85)	-0.64 (1.18)	0.16	119
Dropping 2020–2022	-0.37 (0.83)	-1.08 (1.39)	-0.37 (1.06)	-0.14 (1.46)	0.02	128
Dropping counties first dated after 2024 (leads only)	-0.41 (0.72)	-1.21 (1.22)	-0.62 (0.94)	-0.43 (1.25)	0.05	101
State-by-year fixed effects	-0.11 (0.37)	-0.11 (0.51)	-0.08 (0.35)	0.55 (0.61)	0.03	128
Price in levels (cents per kWh)	0.072 (0.107)	-0.093 (0.112)	-0.001 (0.094)	-0.025 (0.094)	0.03	128
Weighted by residential customers	-0.48 (1.16)	-1.98 (1.16)	-0.83 (1.04)	-0.88 (1.66)	0.62	128
Window -4 to $+6$	-0.41 (0.72)	-1.19 (1.22)	-0.61 (0.94)	-0.42 (1.25)	0.56	128
EPA dates shifted one year earlier	0.27 (0.74)	-0.41 (1.05)	0.08 (0.87)	0.48 (1.18)	0.02	123
EPA dates shifted two years earlier	-0.24 (0.68)	-0.55 (0.84)	-0.40 (0.73)	-0.42 (1.02)	0.01	121
Never-treated controls matched on pre-period level and growth	-0.00 (0.67)	-0.88 (1.12)	-0.61 (0.87)	-0.68 (1.23)	0.07	128
Only counties whose facility is their utility's first, utility-clean controls	0.17 (1.00)	-0.13 (1.63)	0.64 (1.29)	1.39 (1.71)	0.00	64
Wild-cluster bootstrap p (Webb weights): static treated $\times$ post, residential, county and year FE (treated states)	0.639 (38)					

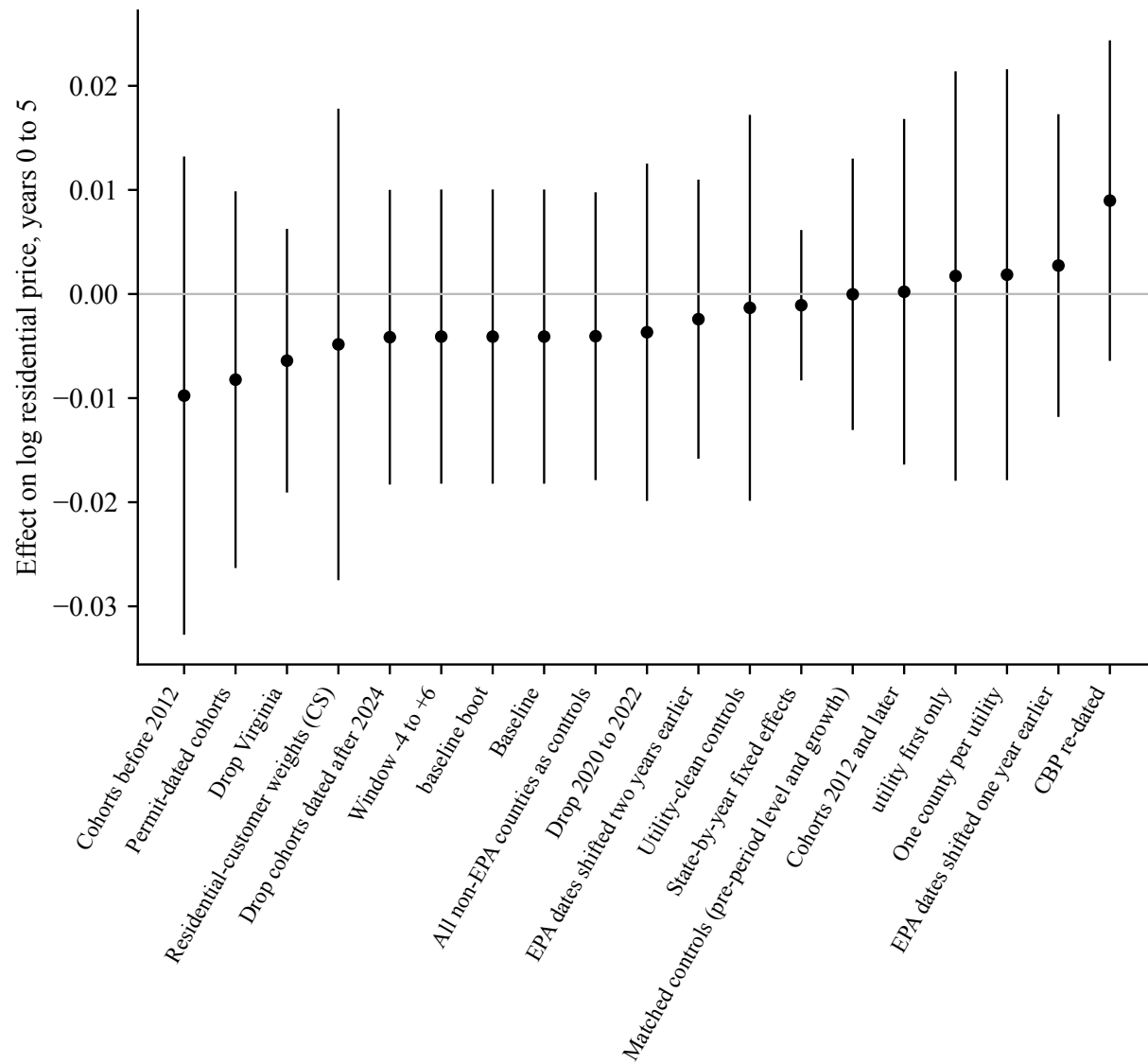
Notes: Each cell is the mean effect over years 0 to 5 in log points with its state-clustered standard error in parentheses. The weighted row uses the Callaway–Sant’Anna estimator, which takes weights. The levels row is in cents per kilowatt-hour. Source: Form EIA-861 and EPA ICIS-Air; authors’ calculations.

matched controls, the date shifts, the dropped future cohorts, and the utility-first events were added after the first results, and are marked as such in the memo), each introduced by the concern it addresses; Figure 7 draws the residential set; Figure 8 draws the leave-one-state-out distribution and the placebo timing.

The control group. Replacing the clean controls with every county without an EPA facility, which adds the counties the map lists and the EPA does not, moves the residential estimate to -0.4 (0.7); if the missed facilities were many and their effects large, including them as controls would have attenuated the estimate, and it did not. Restricting the controls to counties whose largest utility serves no treated county, which removes the within-utility comparisons that are zero by construction, gives -0.1 (0.9); with one treated county per utility as well, so that each utility’s price path is counted once, 0.2 (1.0) on 61 counties.

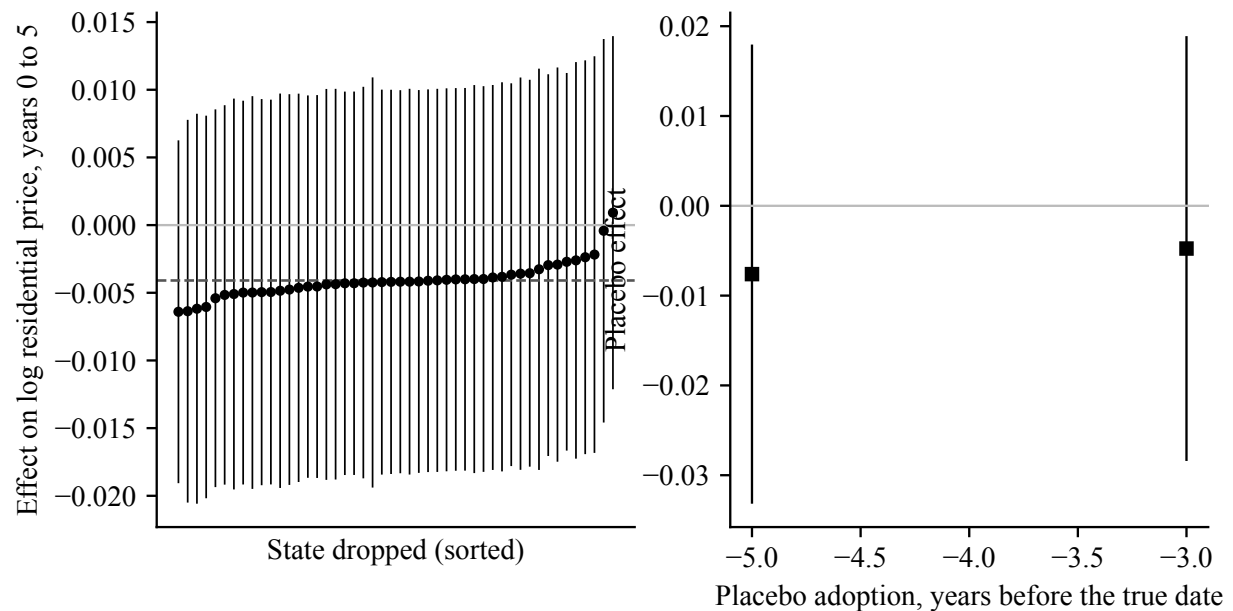
Dates. The permit-dated cohorts were reported in Section 7. Re-dating the openings from the County Business Patterns, the first year a county shows a data-processing establishment with 50 or more employees, gives 0.9 (0.8) on the 50 counties that are not re-dated to the first year of the files. Dropping the 27 counties dated after 2024, whose contribution is pre-opening years in 2019 to 2024, the years in which the interconnection queues filled and anticipation would live, gives -0.4 (0.7). The placebo that moves every opening three and five years earlier, estimated on pre-opening data only, gives -0.5 (1.2) and -0.8 (1.3). Because the known cases say the EPA

Figure 7. The specification curve: the residential estimate across specifications



Notes: Point estimates and 95 percent intervals for the residential mean effect over years 0 to 5 under each specification in Table 7, sorted; the levels specification is omitted because its units differ. Source: Form EIA-861 and EPA ICIS-Air; authors' calculations.

Figure 8. Leave-one-state-out estimates and placebo timing



Notes: Left: the residential mean effect over years 0 to 5 re-estimated dropping each state in turn, sorted, with 95 percent intervals; the dashed line is the full-sample estimate. Right: the static effect when every opening is moved three and five years earlier, estimated on pre-opening observations only. Source: Form EIA-861 and EPA ICIS-Air; authors' calculations.

date is late in most cases and early in one, shifting every opening one year earlier gives 0.3 (0.7) and two years earlier -0.2 (0.7); a late date puts the first years of a true effect into the reference year and the leads, and moving the date back does not uncover one.

Cohorts and samples. The cohorts of 2012 and later give 0.0 (0.8) and the earlier cohorts -1.0 (1.2). Dropping Virginia, the state with the most treated counties, gives -0.6 (0.6); dropping 2020 to 2022, the pandemic and the fuel shock, -0.4 (0.8). Leaving one state out at a time moves the estimate between -0.6 (without Virginia) and 0.1 (without New Jersey) log points; leaving one cohort out moves it between -1.0 (without the 2018 cohort) and -0.1 (without the 2015 cohort); every leave-one-out estimate is within one standard error of the full-sample estimate, and the state range crosses zero because the full-sample estimate is itself within a standard error of zero.

Selection on trajectories, state shocks, form, and weights. Matching each treated county to the ten never-treated counties nearest it in pre-period price level and growth gives -0.0 (0.7). With state-by-year fixed effects, so that the effect is identified from the timing of openings within a state, the residential estimate is -0.1 (0.4). In levels the estimate is 0.072 cents per kilowatt-hour (0.107), imprecise because the Alaska utilities enter in levels. Weighted by

residential customers under Callaway–Sant’Anna, -0.5 (1.2). A narrower window of four years before to six after gives -0.4 (0.7).

Parallel trends. Table C2 reports the Rambachan and Roth (2023) confidence sets for the years 0 to 5 mean under the relative-magnitudes restriction, in which the period-to-period change in the trend violation after the opening is at most $\bar{M}$ times the largest such change before it. At $\bar{M} = 0.5$ the set is -5.0 to 4.5 log points; at $\bar{M} = 1$, -9.1 to 8.7 . On the narrower window of four years before to six after, which drops the year -6 lead, the set at $\bar{M} = 0.5$ narrows to -2.8 to 2.5 . The sets are wide because the year -6 coefficient is the largest lead, and they contain zero at every $\bar{M}$, which for an estimate near zero is not a reassurance but a statement of what the leads allow: a rise of up to five percent is consistent with the data if post-opening trends may diverge by half the largest pre-opening divergence.

The unit, step by step. Table 8 lays out every step from the county baseline toward the price-setting unit: controls served by no treated utility, one treated county per utility, only the events that are the serving utility’s first facility (0.2 on 64 counties, standard error 1.0), investor-owned treated and control counties, and then the utility-by-state unit under its several control pools. The residential estimate moves from a small saving at the county baseline to a rise of one to three log points at the utility unit, and the move is in the control pool: the county rows that remove within-utility comparisons stay near zero, and the utility rows that give the large utilities their weight by class sales agree with the county estimate, while the unweighted utility rows compare large investor-owned utilities with a pool of small municipal and cooperative systems.

The utility unit. Table C3 re-runs the design with the utility-by-state as the unit, the model’s price setter, two ways: every utility that lists a treated county is treated at the county’s first opening, and only the county’s largest utility is. The residential estimates are 1.8 (1.1) for the largest-utility version and 1.0 (0.8) for the territory version, one to two log points above the county estimate and within two standard errors of it. In vertically integrated states, the regime the model’s dividend describes, the largest-utility residential estimate is 2.8 (1.1; wild-cluster bootstrap $p = 0.19$), a rise; against never-treated utilities at least as large as the treated 25th percentile it is 1.1 (1.7) on 6 controls, and against investor-owned never-treated utilities only, 2.0 (1.3) on 68. The county design’s own vertically integrated cell is -0.1 (0.7). The two units disagree on the residential sign in that regime; the disagreement shrinks and loses precision when

Table 8. From the county to the utility: the residential estimate at each step toward the price-setting unit

Design	Residential, years 0–5 (log points)	Pre-trend p	Treated / never
County, clean controls (the headline)	-0.41 (0.72)	0.05	128 / 2,445
County, controls served by no treated utility	-0.13 (0.95)	0.05	128 / 1,157
County, one treated county per utility, utility-clean controls	0.19 (1.01)	0.01	61 / 1,157
County, only events that are the utility’s first, utility-clean controls	0.17 (1.00)	0.00	64 / 1,157
County, investor-owned treated and control counties	-0.61 (0.76)	0.10	116 / 2,016
Utility-by-state, every utility listing a treated county	1.03 (0.78)	0.14	313 / 2,080
Utility-by-state, the county’s largest utility	1.75 (1.07)	0.38	56 / 2,080
controls at least the treated 25th-percentile size	-0.39 (1.87)	0.73	42 / 6
vertically integrated states, investor-owned controls	2.00 (1.28)	0.24	41 / 68
vertically integrated states, size-restricted controls	1.12 (1.71)	0.33	41 / 6
vertically integrated states, all never-treated utilities	2.84 (1.08)	0.07	41 / 1,835

Notes: Sun–Abraham mean effects over years 0 to 5 in log points with state-clustered standard errors; each row is defined in Sections 8 and 9 and in Appendix C. Source: Form EIA-861 and EPA ICIS-Air; authors’ calculations.

the control pool is made like for like, and the class-sales-weighted utility rows, which give the large utilities their weight, agree with the county estimate; the paper reports the unweighted disagreement as open. The industrial estimates differ more: 8.6 (2.4) and 6.6 (1.9), a rise of several percent in the industrial average revenue of treated utilities relative to never-treated ones. Three features of the comparison explain the difference. The treated utilities are large investor-owned utilities, with median residential sales of 4,453,856 megawatt-hours, and the never-treated pool is the small municipal and cooperative utilities, with median sales of 18,669; the comparison is not like for like, and restricting the controls to utilities at least as large as the treated 25th percentile leaves 6 of them, on which the residential estimate is -0.4 (1.9). The industrial series fails its pre-trend test at the utility unit ($p = 0.000$ for the largest-utility version). And the rise is concentrated in restructured states, 20.0 (5.0) against 1.9 (1.3) in vertically integrated ones, where an investor-owned utility’s industrial average revenue is a changing mix of delivery-only and bundled sales as customers shop (Section 3). The county design with one treated county per utility, which weights each utility’s path once as the utility design does, gives an industrial estimate of 1.4 (1.7), so the difference is the control pool and the class’s measurement, not the weighting. The utility-level industrial rise is reported here as the one place the design’s answer depends on the unit, and the data that would settle it, each utility’s load by customer and the class in which it bills each facility, are not public.

9 Mechanisms and heterogeneity

The model predicts where the effect should differ. Table 9 reports the dose-response and Table 10 the splits, each with the inference its size allows: a wild-cluster bootstrap p-value on the cell's static treated-by-post coefficient and the number of states that carry treatment, because state-clustered asymptotic errors on cells carried by three to nineteen treated states are not credible, and the wild-cluster bootstrap under-rejects there too (Section 6).

9.1 *The dose: pass-through per unit of load share*

Proposition 1 says the effect on an incumbent class is the load's share of the serving utility's sales times the incremental-cost premium, so across counties the effect should be proportional to λ_c . Facility size is not in the public record, so the load-share proxy sets every first facility at 100 megawatts and divides by the largest serving utility's state-wide sales in the county's first panel year; it varies across treated counties from 1.0 to 17.6 percent between the 10th and 90th percentiles, around a mean of 6.8 percent (the proxy is measured in the county's first panel year, the calibration's λ in the year before the opening, so the two distributions differ). That is more than an order of magnitude of variation, all of it in the denominator: the proxy ranks counties by how small their utility is, not by how large their facility is, and a true load share with a 2 megawatt facility at a small cooperative and a 200 megawatt facility at a large investor-owned utility can order the two counties the other way. The dose-response is estimated on the stacked clean-control design: the static treated-by-post coefficient reproduces the library implementation of the stacked estimator exactly (the cross-check is in the replication package), and the interaction of treated-by-post with the demeaned proxy is added. In the residential class the effect at the mean proxy is -0.6 log points (1.3) and the slope is -0.02 log points per percentage point of load share (0.08), $p = 0.82$: no gradient, and a slope whose 95 percent interval runs from -0.16 to 0.13 log points per percentage point. Under the model the slope is $\bar{\sigma}\pi/p$, so the interval, divided by the 0.69 of the long-run change in place over the window, allows a premium anywhere between -24 and 19 percent of the retail price; the proxy is too coarse on the numerator for the test to bind. The commercial slope, the class that bills most data centers, is 0.08 (0.10), $p = 0.44$; the industrial slope is -0.23 (0.10), $p = 0.02$, and the gap between the industrial and residential slopes is the composition statistic of Proposition 3 in its natural units: a class that bills the new load shows a steeper decline in the load share than the residential class does, because a fixed-size

Table 9. The dose-response: the effect at the mean load-share proxy and its slope in the proxy, stacked design

Class	Effect at the mean load share (log points)	Slope per percentage point of load share (log points)	<i>p</i> (slope)	Treated counties
Residential	-0.63 (1.27)	-0.017 (0.076)	0.82	128
Commercial	-1.66 (1.72)	0.077 (0.100)	0.44	128
Industrial	-0.98 (1.72)	-0.231 (0.098)	0.02	128
All classes	-0.97 (1.44)	-0.089 (0.077)	0.25	128

Notes: The stacked clean-control design of Cengiz et al. (2019) with the static treated-by-post indicator and its interaction with the demeaned load-share proxy, $\hat{\lambda}_c =$ a 100 MW load at an 80 percent load factor divided by the largest serving utility’s state-wide sales in the county’s first panel year, in percentage points; county-by-stack and year-by-stack fixed effects; standard errors clustered by state. Source: Form EIA-861 and EPA ICIS-Air; authors’ calculations.

customer at a cheap tariff is a larger share of a small utility’s industrial sales. The industrial slope is the one gradient the proxy finds; it has the sign composition predicts, and it is also what a small utility’s industrial average, dominated by a handful of customers, would show mechanically, so the paper does not rest on it.

9.2 Permit class and facility count

The first facility’s Clean Air Act classification is the only size information the public record carries, and it is coarse and not monotone in load: a major source emits enough to need a Title V permit, which for a data center means a large bank of generators, but a large campus can take hour limits on its generators to stay a synthetic minor, and a minor source is small.

Proposition 1 predicts the effect grows in magnitude with size. The residential estimate is -1.9 (0.7; bootstrap $p = 0.41$ on 15 treated states) in the 19 counties whose first facility is a major source, 0.4 (1.1) where it is a synthetic minor, and -0.5 (0.7) where it is a minor source; the commercial estimate in the major-source cell is -1.2 (0.9). Redefining the event as the county’s third dated facility, which the 34 counties that reach one do at a median of several years after the first, gives -1.9 (1.0) in the residential class, -4.7 (1.0) in the industrial class, and -3.1 (0.8) in all classes. Both splits were added after the first results and the Identification Memo records them as additions; they are labeled exploratory in Table 10. The third-facility event has a further problem: counties that go on to a third facility are the ones operators kept choosing, so the event conditions on the outcome path of the first, and its pre-period carries the still-phasing-in effect of the facilities before it. Selection on trajectories threatens the first-facility event too, which is why the matched-control and state-by-year rows of Table 7 are run; they answer it for the first event and cannot for the third, whose pre-period is already treated. It is reported for completeness and

Table 10. Heterogeneity: the effect over years 0 to 5 by class, regime, ownership, utility size, and facility type, with bootstrap inference

Sample	Residential	Commercial	Industrial	All classes	Boot. p (res.)	Treated counties	Treated states
<i>Regulatory structure</i>							
Vertically integrated states	-0.11 (0.66)	0.16 (1.10)	-0.15 (0.97)	-0.24 (0.83)	0.58	89	27
Restructured states	-1.36 (0.82)	-2.88 (1.56)	-0.05 (1.52)	-1.19 (1.16)	0.55	39	11
<i>Ownership of the county's largest utility</i>							
Investor-owned	-0.61 (0.76)	-1.75 (1.26)	-0.77 (1.36)	-0.88 (1.00)	0.41	116	36
Cooperative	1.86 (0.67)	4.58 (1.25)	1.01 (1.63)	-0.32 (1.00)	0.52	4	3
Municipal and public	4.53 (2.49)	4.65 (2.64)	10.38 (2.84)	5.35 (2.78)	0.22	8	3
<i>Size of the serving utility (sales terciles)</i>							
Small	0.41 (1.49)	-0.62 (1.80)	-5.27 (1.79)	-2.15 (1.67)	0.67	25	17
Medium	-0.57 (1.00)	-1.55 (2.13)	1.00 (1.68)	0.11 (1.20)	0.73	44	20
Large	-0.71 (0.92)	-1.29 (1.41)	0.75 (1.81)	-0.46 (1.20)	0.59	59	15
<i>Exploratory: hyperscaler present</i>							
Hyperscaler among first facilities	0.10 (0.83)	0.77 (1.07)	-0.42 (1.19)	-0.35 (0.87)	0.36	30	19
No hyperscaler	-0.56 (0.87)	-1.85 (1.29)	-0.43 (1.48)	-0.70 (1.10)	0.84	98	34
<i>Exploratory: facility size and count</i>							
First facility is a Clean Air Act major source	-1.92 (0.72)	-1.20 (0.86)	-1.29 (1.15)	-2.05 (0.68)	0.41	19	15
First facility is a synthetic minor source	0.39 (1.08)	0.46 (1.42)	-0.03 (1.89)	0.47 (1.30)	0.84	59	26
First facility is a minor source	-0.47 (0.69)	-2.91 (1.08)	0.36 (1.13)	-0.90 (0.62)	0.44	42	19
Event redefined as the county's third facility	-1.89 (0.99)	-4.17 (1.08)	-4.65 (1.02)	-3.07 (0.84)	0.04	34	19

Notes: Sun–Abraham mean effects over years 0 to 5 in log points with state-clustered standard errors; each row restricts the treated counties to the cell named and keeps the full clean control group, except the regime and ownership rows, which restrict both. Regime is the state's (restructured per the APPA 2019 list); ownership is the modal class of the county's largest serving utility over the panel; utility size is the tercile of the largest serving utility's sales. The hyperscaler, facility-class, and third-facility rows are exploratory; the facility-class rows cover the treated counties whose first facility carries a classification, which is not all of them. The bootstrap p-value is a wild-cluster (Webb weights) p-value on the residential cell's static treated-by-post coefficient with county and year fixed effects, clustered by state; treated states is the number of states carrying treatment in the cell. Source: Form EIA-861 and EPA ICIS-Air; authors' calculations.

nothing in the paper rests on it. The major-source cell is the one the paper carries forward as a sign: where the first facility was large, the residential price fell by about two percent on the clustered standard error, and the bootstrap on 15 treated states does not reject zero. The third-facility cell is the one cell whose bootstrap does reject zero, $p = 0.04$, and it is the cell with the selection problem.

9.3 Regime

Section 4 says what the regime split can test: the wholesale term is market-wide and nets out of a within-market comparison, so the testable prediction is the wires-only fixed-cost channel, under which restructured incumbents should see the smaller decline. The residential estimate is -1.4 (0.8) in restructured states and -0.1 (0.7) in vertically integrated ones; the difference is about -1.3 log points with a standard error of 1.1, not distinguishable from zero and, in sign, the

opposite of the prediction. The prediction fails, and the paper records the failure as one. A measurement reason to discount the restructured residential series is also on the record: in a restructured state an investor-owned utility's residential average revenue falls mechanically as residential customers shop, since delivery-only sales enter the denominator at a wires-only price, and the shopping share grew in several of these states over the sample.

9.4 *Ownership and utility size*

Proposition 2's purchased-power case (Section 4) predicts a smaller dividend or a rise for a municipal or cooperative system, which buys its energy and is small, so a facility is a large share of its sales. The residential estimate is -0.6 (0.8) in the 116 counties whose largest utility is investor-owned, a cell that restricts the controls to investor-owned counties as well and so is the like-for-like county comparison the utility-unit design lacks, 1.9 (0.7) in the 4 cooperative counties, and 4.5 (2.5) in the 8 municipal and public ones, where the industrial estimate is 10.4 (2.8). The municipal cell is the one cell in the paper where the load share is largest and the one where prices rose, for every class, with a bootstrap p of 0.22 on 3 treated states; it is 8 counties. Where a data center is a large share of a small purchasing system's sales, the point estimates say a rise, which is the purchased-power case's prediction, and the bootstrap on three treated states cannot distinguish it from zero; an all-class rise with the industrial class largest is also what a re-priced wholesale supply contract or a demand ratchet would produce, and the two cannot be told apart in eight counties. The cooperative cell, 4 counties, is reported and not read. The utility-size terciles, the dose in the other direction, give 0.4 (1.5), -0.6 (1.0), and -0.7 (0.9) for the residential class from small to large, with no gradient, and -5.3 (1.8) for the industrial class in the small tercile, where the composition wedge should be and is largest.

9.5 *Hyperscalers*

An exploratory split by whether a hyperscale operator is among the county's first facilities gives 0.1 (0.8) with and -0.6 (0.9) without; the operator label is lost for the shell companies the hyperscalers permit under, so the split is noisy on both sides and the paper draws nothing from it.

Taken together, the splits say that the effect does not move detectably with the serving utility's size; that its point estimates are negative where the first facility was a major source and positive in the small public systems that purchase their power, each with the model's sign and

neither distinguishable from zero once the inference fits the handful of treated states; and that the wires-channel prediction for restructured states fails in sign.

10 Implications

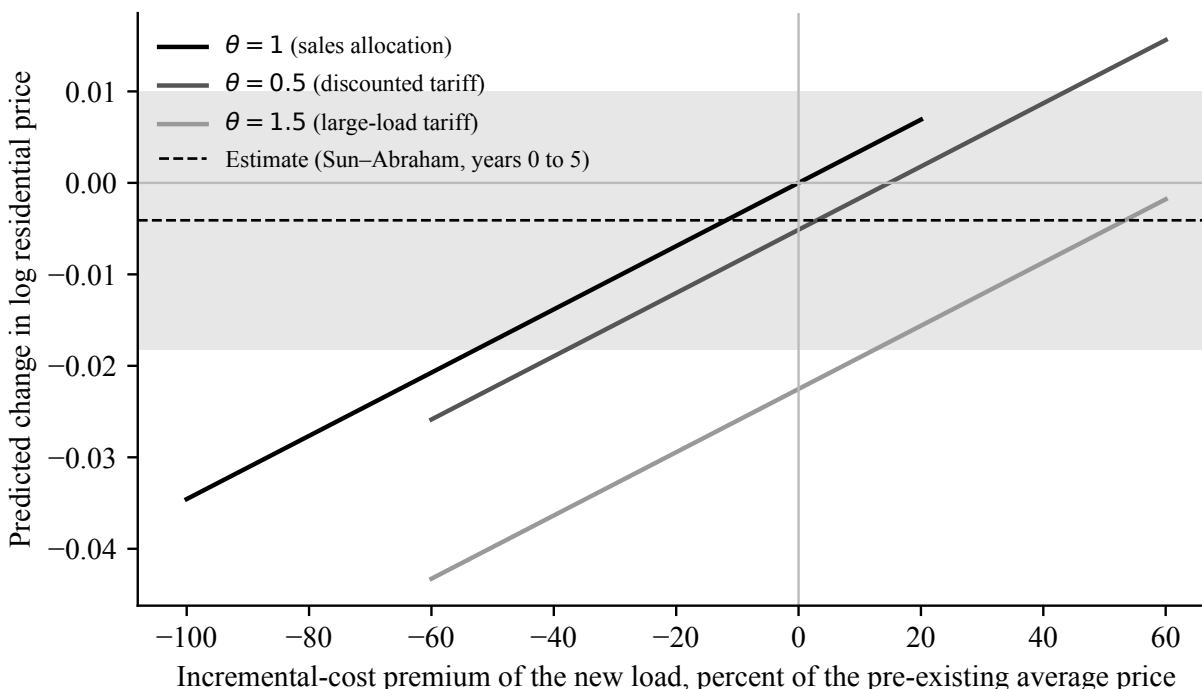
10.1 *The estimate against the calibrated model*

Figure 9 draws the model's predicted change in the residential price over the incremental-cost premium, at the median load share, for three allocation intensities, with the estimate and its 95 percent band overlaid. The band cuts the sales-allocation line at a premium between about -76 and 42 percent of the retail price at a 100 MW load; a new load that cost the system much less than it paid, the slack case, is to the left of the band, and a load on a steep discount in a tight system is to the right. The slack case the parallel-trends interval excludes is itself the least plausible one under merit-order dispatch, where the marginal unit's cost is at or above the average unless embedded cost carries above-market legacy contracts, so the exclusion is of a case that mostly does not arise. The width of the range is the load assumption: at 30 MW the same band implies -248 to 136 percent, at 10 MW -738 to 407 , and at 300 MW -27 to 15 ; under the weaker trend assumption the 100 MW range is -209 to 190 percent. At the 90th percentile of the load share, where the same premium produces an effect four times larger, the parallel-trends range is -18 to 10 percent of price. The first stage says the typical first opening was nearer the small end of these loads, so the data through 2024 do not bound the premium of the typical opening and say nothing about a premium at the loads now being built.

10.2 *Virginia*

The state with the largest concentration of data centers is the natural aggregate check, because the county design cannot see a utility-wide change that reaches every county a utility serves. Figure 10 draws Virginia's log residential price against a synthetic Virginia built from the vertically integrated states, with the outcome's own lags as predictors, treated in 2015, the year the hyperscale build-out in Loudoun and the surrounding counties accelerated. The post-2015 gap averages -2.4 log points, with a pre-period root mean squared error of 0.015 and a ratio of post- to pre-period error of 2.0; against the in-space placebos, 29 donor states each treated in turn, the permutation p-value is 0.57. The exercise has low power, since several donors have larger ratios from their own rate events, and it is reported as consistent with the county estimates rather

Figure 9. The calibrated model’s predicted residential price change over the incremental-cost premium, with the estimate



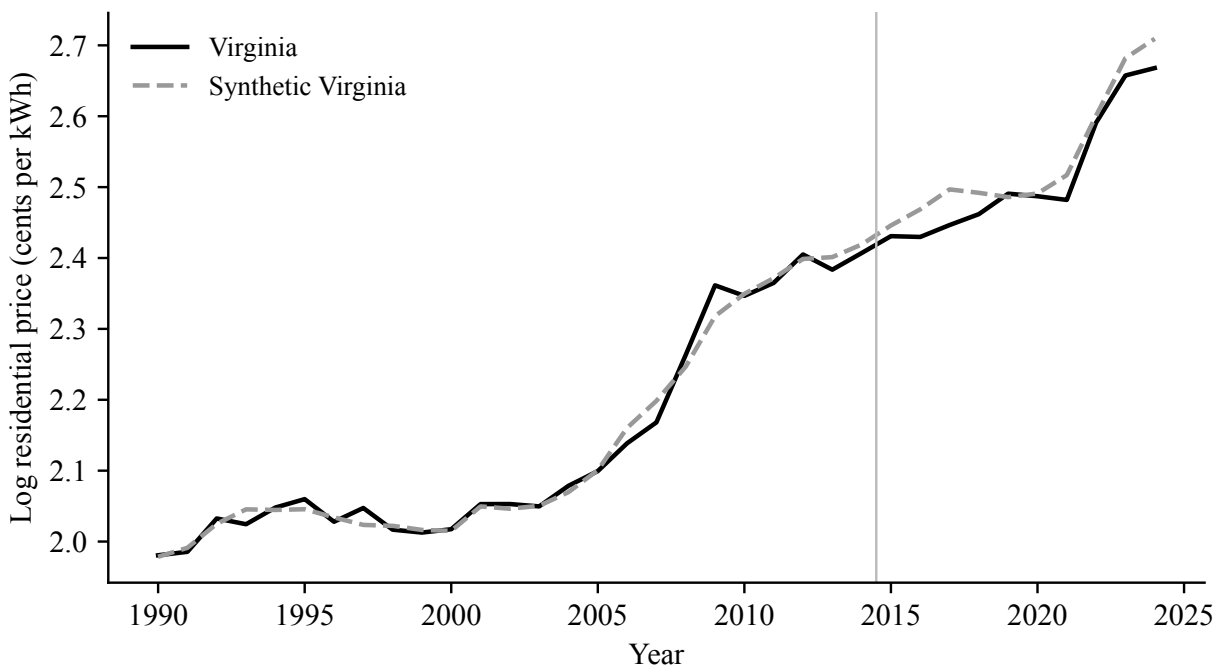
Notes: Each line is the model’s $\Delta \log p$ for an incumbent class at the median load share over treated counties, as the incremental-cost premium of the new load varies, for a sales allocator ($\theta = 1$), a discounted tariff ($\theta = 0.5$), and a large-load tariff ($\theta = 1.5$); the dashed line and shaded band are the Sun–Abraham residential estimate over years 0 to 5 and its 95 percent interval. Source: Table 1; authors’ calculations.

than as a confirmation. A version treated in 2008, the year Virginia’s tenth treated county was dated, shows a positive gap in 2008 to 2012 that coincides with the first rate-assessment window, 2007 to 2012, of Virginia’s 2007 Electric Utility Regulation Act (Virginia State Corporation Commission, 2023), and a negative one after 2013; its average is -0.4 log points with $p = 0.52$, and the paper reads that date as confounded.

10.3 What the result means for rate design, and what it does not

The model says the sign of the pass-through is set by three things a commission controls: whether the new load pays the class tariff or a discount, how much of its network cost it contributes up front, and how its share of fixed cost is allocated. The large-load tariffs that E3 (2026) count, with minimum demand charges and contributions in aid of construction, raise θ and ϕ and move the sign toward the dividend. Under the tariffs of 2002 to 2024 the typical opening did not raise the residential price by more than a few dollars a month (Section 7);

Figure 10. Virginia’s residential price and its synthetic control



Notes: Log residential price for Virginia and for the synthetic Virginia built from vertically integrated donor states on the outcome’s pre-period lags, treated in 2015 (the vertical line); the 2008 version is reported in the text. Source: Form EIA-861; authors’ calculations.

whether it paid its own incremental cost the data cannot say, and the point estimate is a small saving the data cannot distinguish from zero.

What the result does not say is what the facilities now under construction will do. The 27 counties whose first facility the EPA list dates to 2025 or later have no post-opening year in the price panel, and the facilities of that wave are an order of magnitude larger than the 100 megawatt calibration: a load of a gigawatt at the median serving utility is a load share above the sample’s 90th percentile, at which the tight-system, discounted-tariff case predicts a rise of 5.2 log points and the neutral sales-allocation case a fall of 5.7. The design can be re-run as those years arrive, and the first stage is the statistic to watch: when the serving utility’s sales move at an opening, the price regression has a treatment it can see. The premium is also not a parameter of nature: data centers located, over these years, where power was cheap and capacity slack, which is where the premium was lowest, so the sample’s premium is selected downward, and the loads now being connected exhaust the slack the sample was selected on. The data that would decide the question sooner are not public: each utility’s load by customer, the class in which it bills each facility, and the terms of its special contracts, which Martin and Peskoe (2025) show are often confidential by

design.

11 Conclusion

When a county received its first data center between 2002 and 2024, the retail electricity price its households paid did not move by more than the design can see: the estimate is -0.4 log points over the six years after the opening, with a 95 percent set of -5.0 to 4.5 under modest violations of parallel trends and -1.8 to 1.0 under none, which on the typical bill is a few dollars a month at most. Three estimators that handle staggered adoption agree with each other and with the benchmark, and the serving utilities' sales did not detectably move at the opening, so the typical first facility was a small load. A model of cost-of-service rate design says the effect scales with the load's share of the serving utility's sales times its incremental-cost premium, and the design bounds that premium only on an assumption about the load. Where the first facility was a Clean Air Act major source the residential price fell by about two percent and in the small public systems that purchase their power it rose, the signs the model predicts, but neither cell survives inference that fits a handful of treated states; and a load-share proxy that can vary only with the utility's size shows no residential gradient.

The design cannot say three things. The facility list is built from a federal air register that some states do not fill, and its dates run late, so the county estimates are attenuated by an amount the known-cases test bounds but does not remove. The county price is a utility average, and the design re-run at the utility unit gives, in vertically integrated states, a residential rise that the county design does not show and that shrinks and loses precision when the control pool is made like for like; its industrial series rises more, for a measurement reason the paper names, the mix of delivery-only and bundled sales in restructured states, without resolving it. And the wave of facilities dated 2025 and later, an order of magnitude larger than those in the sample, has no post-opening year yet.

The open question is the dose. The model says the pass-through to incumbents is the load share times the incremental-cost premium, and the data through 2024 cover first facilities too small to move their utility's sales. At the loads now being connected the same premium produces effects large enough to measure in a single rate case, and the tariffs under which they are connected set its sign. The next vintage of these files will carry the first post-opening years of that wave, and the specification in this paper is written to be re-run on it.

References

- Abadie, Alberto**, “Using Synthetic Controls: Feasibility, Data Requirements, and Methodological Aspects,” *Journal of Economic Literature*, 2021, 59 (2), 391–425.
- , **Alexis Diamond**, and **Jens Hainmueller**, “Synthetic Control Methods for Comparative Case Studies: Estimating the Effect of California’s Tobacco Control Program,” NBER Working Paper 12831, National Bureau of Economic Research 2010. Published in *Journal of the American Statistical Association* 105(490), 2010.
- Alvarez, Fernando**, **David Argente**, **Joyce Chow**, and **Diana Van Patten**, “Data Centers and Local Economies in the Age of AI: A Shift-Share Approach,” NBER Working Paper 35194, National Bureau of Economic Research 2026.
- American Public Power Association**, “Retail Electric Rates in Deregulated and Regulated States: 2018 Update,” Technical Report, American Public Power Association 2019.
- Baker, Andrew C.**, **David F. Larcker**, and **Charles C. Y. Wang**, “How Much Should We Trust Staggered Difference-in-Differences Estimates?,” Finance Working Paper 736/2021, European Corporate Governance Institute 2022. Published in *Journal of Financial Economics* 144(2), 2022.
- Bertrand, Marianne**, **Esther Duflo**, and **Sendhil Mullainathan**, “How Much Should We Trust Differences-in-Differences Estimates?,” NBER Working Paper 8841, National Bureau of Economic Research 2004. Published in *Quarterly Journal of Economics* 119(1), 2004.
- Borenstein, Severin**, “The Economics of Fixed Cost Recovery by Utilities,” *The Electricity Journal*, 2016, 29 (7), 5–12.
- and **James B. Bushnell**, “Do Two Electricity Pricing Wrongs Make a Right? Cost Recovery, Externalities, and Efficiency,” NBER Working Paper 24756, National Bureau of Economic Research 2022. Published in *American Economic Journal: Economic Policy* 14(4), 2022.
- and **James Bushnell**, “The U.S. Electricity Industry after 20 Years of Restructuring,” NBER Working Paper 21113, National Bureau of Economic Research 2015. Published in *Annual Review of Economics* 7, 2015.

Borusyak, Kirill, Xavier Jaravel, and Jann Spiess, “Revisiting Event-Study Designs: Robust and Efficient Estimation,” *Review of Economic Studies*, 2024, 91 (6), 3253–3285.

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, Jonah B. Gelbach, and Douglas L. Miller, “Bootstrap-Based Improvements for Inference with Clustered Errors,” NBER Technical Working Paper 344, National Bureau of Economic Research 2008. Published in *Review of Economics and Statistics* 90(3), 2008.

Cengiz, Doruk, Arindrajit Dube, Attila Lindner, and Ben Zipperer, “The Effect of Minimum Wages on Low-Wage Jobs: Evidence from the United States Using a Bunching Estimator,” NBER Working Paper 25434, National Bureau of Economic Research 2019. Published in *Quarterly Journal of Economics* 134(3), 2019.

Cicala, Steve, “When Does Regulation Distort Costs? Lessons from Fuel Procurement in U.S. Electricity Generation,” NBER Working Paper 20109, National Bureau of Economic Research 2015. Published in *American Economic Review* 105(1), 2015.

—, “Imperfect Markets versus Imperfect Regulation in U.S. Electricity Generation,” NBER Working Paper 23053, National Bureau of Economic Research 2022. Published in *American Economic Review* 112(2), 2022.

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.

— **and** —, “Two-Way Fixed Effects and Differences-in-Differences with Heterogeneous Treatment Effects: A Survey,” *The Econometrics Journal*, 2023, 26 (3), C1–C30.

E3, “Understanding the Drivers of Rising Electricity Rates and the Role of Data Centers,” Technical Report, Energy and Environmental Economics, Inc. 2026.

Electric Power Research Institute, “Powering Intelligence: Analyzing Artificial Intelligence and Data Center Energy Consumption,” Technical Report, Electric Power Research Institute 2024.

Fabra, Natalia and Mar Reguant, “Pass-Through of Emissions Costs in Electricity Markets,” NBER Working Paper 19613, National Bureau of Economic Research 2014. Published in *American Economic Review* 104(9), 2014.

Fabrizio, Kira R., Nancy L. Rose, and Catherine D. Wolfram, “Do Markets Reduce Costs? Assessing the Impact of Regulatory Restructuring on US Electric Generation Efficiency,” NBER Working Paper 11001, National Bureau of Economic Research 2007. Published in *American Economic Review* 97(4), 2007; the working paper lists the first author as Kira Markiewicz.

Forrester, Sydney, Andrew Satchwell, Galen Barbose, Evan Cappers, Cesca Miller, and Andrew Alberg, “Retail Electricity Price and Cost Trends: 2024 Update,” Technical Report, Lawrence Berkeley National Laboratory 2024.

Freyaldenhoven, Simon, Christian Hansen, and Jesse M. Shapiro, “Pre-Event Trends in the Panel Event-Study Design,” NBER Working Paper 24565, National Bureau of Economic Research 2019. Published in *American Economic Review* 109(9), 2019.

Ganapati, Sharat, Joseph S. Shapiro, and Reed Walker, “Energy Cost Pass-Through in U.S. Manufacturing: Estimates and Implications for Carbon Taxes,” NBER Working Paper 22281, National Bureau of Economic Research 2020. Published in *American Economic Journal: Applied Economics* 12(2), 2020.

Gardner, John, “Two-Stage Differences in Differences,” *arXiv preprint 2207.05943*, 2022.

Garimella, Kiran, “Do Data Centers Raise Residential Electricity Prices? Evidence from 24 U.S. States,” Preliminary draft, Rutgers University 2026.

Golden, Dana, Aruna Balasubramanian, and Niranjan Balasubramanian, “Certificates without Electrons? Theory and Evidence on Impacts from AI-Driven Power Demand,” *arXiv preprint 2606.00811*, 2026.

Goodman-Bacon, Andrew, “Difference-in-Differences with Variation in Treatment Timing,” NBER Working Paper 25018, National Bureau of Economic Research 2021. Published in *Journal of Econometrics* 225(2), 2021.

International Energy Agency, “Energy and AI,” World Energy Outlook Special Report, International Energy Agency 2025.

Joint Legislative Audit and Review Commission, “Data Centers in Virginia,” Report to the Governor and the General Assembly of Virginia 598, Joint Legislative Audit and Review Commission, Commonwealth of Virginia, Richmond, VA December 2024.

Joskow, Paul L., “Inflation and Environmental Concern: Structural Change in the Process of Public Utility Price Regulation,” *Journal of Law and Economics*, 1974, 17 (2), 291–327.

—, “Incentive Regulation and Its Application to Electricity Networks,” *Review of Network Economics*, 2008, 7 (4), 547–560.

— **and Richard Schmalensee**, “Incentive Regulation for Electric Utilities,” *Yale Journal on Regulation*, 1986, 4 (1), 1–49.

— **and —**, “Cost of Service Regulation of Electricity Distribution Services in the U.S.,” in “Handbook chapter, March 26, 2024 draft,” Massachusetts Institute of Technology, 2024.

Kay, Owen, Robert Reaser, and Reid Taylor, “Processing Power: The Effect of Data Centers on Wholesale Electricity Markets,” Working Paper 2606, Federal Reserve Bank of Dallas 2026.

Knittel, Christopher R., Juan Ramon L. Senga, and Shen Wang, “Flexible Data Centers and the Grid: Lower Costs, Higher Emissions?,” NBER Working Paper 34065, National Bureau of Economic Research 2025.

Lazar, Jim and Wilson Gonzalez, “Smart Rate Design for a Smart Future,” Technical Report, Regulatory Assistance Project 2015.

—, **Paul Chernick, and William Marcus**, “Electric Cost Allocation for a New Era: A Manual,” Technical Report, Regulatory Assistance Project 2020. Edited by Mark LeBel.

- MacKinnon, James G. and Matthew D. Webb**, “The Wild Bootstrap for Few (Treated) Clusters,” QED Working Paper 1364, Queen’s University, Department of Economics 2018. Published in *The Econometrics Journal* 21(2), 2018.
- Martin, Eliza and Ari Peskoe**, “Extracting Profits from the Public: How Utility Ratepayers Are Paying for Big Tech’s Power,” Technical Report, Harvard Law School Electricity Law Initiative 2025.
- MIT Climate Policy Center**, “Data Centers and the Grid,” Policy brief, MIT Climate Policy Center July 2025. Policy brief; its source line credits Knittel, C.R., Senga, J.R.L., and Wang, S. (2025), MIT CEEPR WP 2025-14.
- MIT Energy Initiative**, “Utility of the Future: An MIT Energy Initiative Response to an Industry in Transition,” Technical Report, Massachusetts Institute of Technology, Cambridge, MA 2016. In collaboration with IIT-Comillas.
- Muller, Nicholas Z.**, “Measuring the Impact of Data Centers in the United States Economy: Monetary Damage from Air Pollution and Greenhouse Gas Emissions,” NBER Working Paper 35100, National Bureau of Economic Research 2026.
- National Renewable Energy Laboratory**, “An Introduction to Retail Electricity Choice in the United States,” Technical Report NREL/BR-6A50-68993, National Renewable Energy Laboratory 2017.
- Rambachan, Ashesh and Jonathan Roth**, “A More Credible Approach to Parallel Trends,” *Review of Economic Studies*, 2023, 90 (5), 2555–2591.
- Roth, Jonathan**, “Pretest with Caution: Event-Study Estimates after Testing for Parallel Trends,” *American Economic Review: Insights*, 2022, 4 (3), 305–322.
- , **Pedro H. C. Sant’Anna, Alyssa Bilinski, and John Poe**, “What’s Trending in Difference-in-Differences? A Synthesis of the Recent Econometrics Literature,” *Journal of Econometrics*, 2023, 235 (2), 2218–2244.
- Scalera, Cameron, Valentina Bosetti, and Filippo Pecci**, “Large-Load Electricity Customers and Spillovers on Household Prices: Evidence from U.S. Data Centers,” CEPR Discussion Paper 21854, Centre for Economic Policy Research 2026.

Shehabi, Arman, Sarah J. Smith, Alex Hubbard, Alex Newkirk, Nuoa Lei, Md Abu Bakar Siddik, Billie Holecek, Jonathan Koomey, Eric Masanet, and Dale Sartor, “2024 United States Data Center Energy Usage Report,” Technical Report, Lawrence Berkeley National Laboratory 2024.

—, **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.

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

Virginia State Corporation Commission, “Status Report: Implementation of the Virginia Electric Utility Regulation Act,” Technical Report, Virginia State Corporation Commission, Richmond, VA November 2023. Report to the Governor, the General Assembly and the Commission on Electric Utility Regulation pursuant to § 56-596 B of the Code of Virginia.

—, “SCC Issues Order in Dominion Energy Virginia Biennial Review Case (PUR-2025-00058),” News release, Richmond, VA 2025.

Watten, Asa, John Bistline, and Geoffrey Blanford, “Have Data Centers Raised Your Electric Bill? Causal Evidence from the United States,” *arXiv preprint 2606.19777*, 2026.

Wiser, Ryan, Eric O’Shaughnessy, Galen Barbose, Peter Cappers, and Will Gorman, “Factors Influencing Recent Trends in Retail Electricity Prices in the United States,” *The Electricity Journal*, 2025, 38 (4).

—, **Galen Barbose, Will Gorman, Eric O’Shaughnessy, Sydney Forrester, Paul Donohoo-Vallett, Peter Cappers, Jeffrey Deason, Ryan Hledik, and Long Lam**, “Retail Electricity Price Trends and Drivers: 2026 Edition,” Technical Report, Lawrence Berkeley National Laboratory and The Brattle Group 2026.

Wolak, Frank A., “Public Utility Pricing and Finance,” in Steven N. Durlauf and Lawrence E. Blume, eds., *The New Palgrave Dictionary of Economics*, 2nd ed., Palgrave Macmillan, 2008.

A Proofs and the general allocator

A.1 The general allocator

Every statement in this appendix is a claim checked symbolically or by sampling the assumption domain in the replication package's model file; the identifiers in parentheses name the claims. Let the embedded-cost study allocate the fixed cost K to incumbent classes by shares σ_j with $\sum_j \sigma_j = 1$, so that class j pays $p_j = c + \sigma_j K / q_j$ before the new load. After the new load the regulator sets $c' = (cQ + mL) / (Q + L)$, $K' = K + (1 - \phi)\gamma L$, allocates the share $\theta\lambda$ of K' to the new load and the remainder to the incumbents in their pre-existing proportions, so class j pays $p'_j = c' + \sigma_j(1 - \theta\lambda)K' / q_j$. Writing $k_j = \sigma_j K / q_j$ for class j 's embedded fixed cost per megawatt-hour (a level, the class analogue of k ; the share $\kappa = k/p$ of the main text is a different object) and $\rho_j = \sigma_j Q / q_j$ for the ratio of its allocated share to its sales share,

$$\Delta p_j = \lambda(-c + m) - k_j \lambda \theta + \gamma \rho_j (1 - \phi)(-\lambda \theta + 1) \frac{\lambda}{1 - \lambda} \quad (14)$$

which is equation (5) with k replaced by the class's own embedded fixed cost and the network-cost term scaled by ρ_j . A class allocated more than its sales share ($\rho_j > 1$, a peaky residential class under a demand allocator) bears more of the new load's uncontributed network cost and takes over more fixed cost when the new load pays; the signs of every derivative in Proposition 2 are unchanged, and Proposition 1 holds class by class with $\pi_j = m + \rho_j(1 - \phi)\gamma - c - k_j$ under $\theta = 1$; each statement is a checked claim (*A_general_**), and the main text's equation (5) is the case $k_j = k$, $\rho_j = 1$. The main text's sales allocator is the case $\sigma_j = q_j / Q$, $\rho_j = 1$, $k_j = k$.

A.2 Proof of Proposition 1

From equations (1) to (4), $p'_j - p_j = (c' - c) + (1 - \theta\lambda)K' / Q - k$. The first term is $\lambda(m - c)$ by equation (2) and the definition of λ . The second, by equation (3) and $L/Q = \lambda / (1 - \lambda)$, is $(1 - \theta\lambda)[k + (1 - \phi)\gamma\lambda / (1 - \lambda)]$. Collecting terms gives equation (5); at $\theta = 1$ the fixed-cost terms are $-\lambda k + (1 - \phi)\gamma\lambda$, so $\Delta p_j = \lambda[m - c + (1 - \phi)\gamma - k] = \lambda\pi$, equation (7), and $\partial \Delta p_j / \partial \lambda = \pi$. The sign follows. Every step was checked symbolically (*model/model.py*, claims *PI_** and *display_dp_equals_checked*).

A.3 Proof of Proposition 2

Differentiating equation (5) with respect to the three rate-design parameters gives

$$\frac{\partial \Delta p_j}{\partial \theta} = -\frac{\lambda (\gamma \lambda (\phi - 1) + k (\lambda - 1))}{\lambda - 1} \quad (15)$$

$$\frac{\partial \Delta p_j}{\partial \phi} = \frac{\gamma \lambda (-\lambda \theta + 1)}{\lambda - 1} \quad (16)$$

$$\frac{\partial \Delta p_j}{\partial \gamma} = -\frac{\lambda (\phi - 1) (\lambda \theta - 1)}{\lambda - 1} \quad (17)$$

With $\lambda \in (0, 1)$, $\phi \in [0, 1)$ and $\theta \lambda < 1$, the first is negative, the second is negative and the third is positive, which are parts (i) to (iii). In the restructured case the energy term is $\eta L \geq 0$ in place of $\lambda(m - c)$ and the fixed-cost terms are unchanged with the wires values, equation (8); under slack, $m = c - \Delta_{cm}$ with $\Delta_{cm} > 0$, the difference between the restructured and the regulated energy channel is

$$\Delta p_j^{R,E} - \Delta p_j^E = \Delta_{cm} \lambda + L \eta \quad (18)$$

which is positive. Every derivative and the regime gap were checked by sampling the assumption domain (claims P2_*)).

A.4 Proof of Proposition 3

With the new load billed in class I at p_D and $\mu = L/(q_I + L)$, the class's measured average revenue per megawatt-hour after the opening is $(1 - \mu)p'_I + \mu p_D = p'_I - \mu(p'_I - p_D)$, so the measured change is $\Delta p_I - \mu \delta$ with $\delta = p'_I - p_D$ (equation (9)), and the wedge $-\mu \delta$ is negative for $\delta > 0$ with derivative $-\delta$ in μ (claims P3_*)).

A.5 Proof of Proposition 4

With a constant hazard h the probability that no rate case has occurred by τ years after the opening, counting the opening year, is $(1 - h)^{\tau+1}$, so the share of the long-run change in place is $s_\tau = 1 - (1 - h)^{\tau+1}$ and $\beta_\tau = s_\tau \Delta p_j$. Then $(\beta_{\tau+1} - \beta_\tau)/\Delta p_j = h(1 - h)^{\tau+1} > 0$ (equation (12)), which is positive and decreasing in τ , $s_\tau < 1$, and $1 - s_\tau = (1 - h)^{\tau+1} \rightarrow 0$ (claims P4_*)).

A.6 *Incumbent demand response*

If incumbent sales fall by ε percent for each percent rise in their price, the revenue requirement is recovered over fewer megawatt-hours and the per-megawatt-hour fixed charge rises to first order by

$$\Delta p_j^\varepsilon = \varepsilon k \frac{\Delta p_j}{c + k} \quad (19)$$

a term proportional to Δp_j and scaled by $\varepsilon k/p$, a few tenths at most, that carries its sign and does not change it (claims *A_demand_response_**); for the magnitudes in Table 1 and a retail price elasticity of a few tenths it is below a tenth of a log point.

B *Data construction*

B.1 *The EIA-861 arithmetic*

For every utility u , state s and year t , Form EIA-861 reports revenue R , sales S , and customers N by class j and by part: bundled service (Part A), energy-only service (Part B), delivery-only service (Part C), and bundled service provided by an adjustment entity (Part D). The utility-state class price is $p_{usjt} = (R^A + R^B + R^C + R^D)/(S^A + S^C + S^D)$ in cents per kilowatt-hour, with energy-only sales excluded from the denominator so that a kilowatt-hour delivered by one utility and sold by another is counted once, which is the Administration's own rule for its published state averages. Before 1999 the form had no part column and all service was bundled; the transportation class replaced the "other" class in 2003; Part D begins in 2005; ownership is in the sales file from 2001 and is joined from the utility file for 1990 to 2000. From 2012 the short form reports totals only; 1,124 utilities filed it in 2012 and 1,724 in 2024, and in 2019, a year in which the short-form respondents also filed the long form, they appear inside the sales file, which the construction keeps disjoint from the short-form population in every year. The county price is $p_{cjt} = \sum_{u \in U(c)} w_{ujt} p_{usjt}$ with $w_{ujt} = S_{usjt} / \sum_{u' \in U(c)} S_{u'sjt}$ and $U(c)$ the utilities listing county c in state s that year. Over 2012 to 2024 the construction reproduces the county series in the authors' data library on every cell to within 10^{-13} cents.

B.2 *The county crosswalk*

The utilities' county strings are matched to Census county codes by exact name within state, then by rule: the Virginia independent cities written bare, 30 recurring misspellings each

checked against the state’s county list, counties that changed codes or names (Dade to Miami-Dade; Shannon to Oglala Lakota; Wade Hampton to Kusiavak; Bedford City merged into Bedford County in 2013) kept at the code of the year listed; 66 known cases are asserted in the code, including four that must not match. Unmatched on purpose are 415 listings (68 distinct strings): Alaska villages that are not boroughs, the form’s placeholders, multi-county strings from three utilities in 2003 to 2010, and listings under the wrong state. In Texas the competitive retailers file no territory, so the territory file covers 0.994 of the state’s sales in 2001 and 0.340 in 2024; the paper drops Texas. The next-lowest coverage in 2024 is South Dakota at 0.831.

B.3 The facility list

ICIS-Air holds 280,208 facilities nationally. The inclusion rule, by tier: 504 facilities carry NAICS 518210 in ICIS-Air or, for the same registry identifier, in the Facility Registry Service’s cross-program industry codes; 62 are named data center (in any of the spellings the records use), data services, or colocation, or carry an Amazon Web Services site code beside a Vadata or Amazon Data Services name; 57 carry the bare name of a pure-play colocation operator or a hyperscale operator. Telecommunications switching sites (NAICS 517 without 518210, or a mobile switching office, cell site, or central office in the name) are excluded unless the name says data center; two were excluded. Three traps the rule had to learn: “COLO” is Colorado’s abbreviation in that state’s records; “AMAZON.COM SERVICES” sites with warehouse codes are fulfillment centers; and seven NAICS 518210 facilities at gas wells are cryptocurrency miners, kept and flagged. Of the 623 facilities, 126 are hyperscale operators’ (Amazon, Microsoft, Meta, Google, Oracle, Apple), 19 of them identified through press-documented shell names, which set the operator label only and never trigger inclusion. County codes come from the registry’s air-program link for 597 facilities and the ECHO exporter for 21; 5 have none.

Dating: each air program at a facility carries a begin date and each compliance evaluation a date. A begin date shared by 2,000 or more program rows nationally is a bulk-load stamp; there are 17, led by October 19, 2014 (213,627 rows), the ICIS-Air migration. A facility is dated by the earliest informative program begin date (“permit”), else by its first compliance evaluation (“evaluation”), provided that date precedes its earliest stamp; otherwise it is censored with the stamp as an upper bound. By facility: 270 permit, 248 evaluation, 56 censored at the 2014

migration, 38 censored at a later state reload, 11 with no record. 24 facilities are planned or under construction and 51 permanently closed; both enter the county's first-opening date if dated, since the question is when a county first had one.

B.4 The known-cases test

Table B1 lists the 31 documented hyperscale openings, from the operators' own published site lists, against what the EPA list finds in the same county under the same operator. 15 are found; of the 14 with a usable date, 4 are inside the documented in-service year, 7 are one to three years late, 2 more than three years late, and 1 early; the mean gap is 1.2 years and the median 1. The documented years are the operators' published in-service years, not ground-breaking or announcement dates. Of the 16 misses, 9 counties hold a data-center facility under another name and 7 hold none.

B.5 Universes for the control group

The Global Data Center Map lists 8,818 U.S. facilities with addresses and no dates; the 3,171 with a five-digit ZIP map to 433 counties through the 2020 ZCTA-to-county relationship file. 74 percent of the EPA counties appear in the map, and 1,243 map facilities sit in counties with no EPA facility, mostly carrier hotels and small colocation sites in large metropolitan counties whose generators are permitted locally; those counties are excluded from the clean control group. The IM3 Open Source Data Center Atlas (PNNL) could not be downloaded by script and is not used. The County Business Patterns files carry the industry under SIC 7374 (1986 to 1997), NAICS 514210 (1998 to 2002) and NAICS 518210 (2003 to 2023); establishment counts and size classes are exact through 2016, after which the Bureau's disclosure method hides cells with fewer than three establishments, so the re-dating uses the first year employment is certainly at or above 50.

C Additional results

Figure C1 draws the all-classes event study, Table C1 the Goodman-Bacon decomposition, Table C2 the Rambachan–Roth sets, and Table C3 the utility-unit design; the body refers to each where it is used.

Table B1. The known-cases test: documented hyperscale openings against the EPA list

Operator	Site (county, state)	Documented opening	Result	EPA date	Date quality	Gap (years)
Google	The Dalles, Wasco OR	2006	MISS	—	—	—
Google	Council Bluffs, Pottawattamie IA	2009	MISS	—	—	—
Google	Lenoir, Caldwell NC	2008	MISS	—	—	—
Google	Moncks Corner, Berkeley SC	2007	HIT	2013-05-03	evaluation	6
Google	Pryor Creek, Mayes OK	2007–2012	MISS	—	—	—
Google	Douglas County, Douglas GA	2003	HIT	2003-09-30	evaluation	0
Meta/Facebook	Prineville, Crook OR	2011	MISS	—	—	—
Meta/Facebook	Forest City, Rutherford NC	2012	MISS	—	—	—
Meta/Facebook	Altoona, Polk IA	2014	MISS	—	—	—
Meta/Facebook	Fort Worth, Tarrant TX	2017	MISS	—	—	—
Meta/Facebook	Los Lunas, Valencia NM	2016–2019	HIT	2024-06-18	permit	5
Meta/Facebook	Sarpy, Sarpy NE	2017–2020	HIT	2020-11-19	permit	0
Meta/Facebook	New Albany, Licking OH	2017–2019	MISS	—	—	—
Meta/Facebook	Henrico, Henrico VA	2020	HIT	2021-09-27	permit	1
Meta/Facebook	Huntsville, Madison AL	2021	HIT	2018-03-26	permit	-3
Meta/Facebook	Stanton Springs, Newton GA	2018–2020	HIT	2022-06-30	permit	2
Meta/Facebook	Eagle Mountain, Utah UT	2021	MISS	—	—	—
Meta/Facebook	DeKalb, DeKalb IL	2020–2022	MISS	—	—	—
Meta/Facebook	Gallatin, Sumner TN	2024	MISS	—	—	—
Microsoft	Quincy, Grant WA	2007	MISS	—	—	—
Microsoft	San Antonio, Bexar TX	2008	HIT	—	censored_bulk	—
Microsoft	Boydton, Mecklenburg VA	2010	HIT	2010-09-14	evaluation	0
Microsoft	West Des Moines, Dallas/Polk IA	2012	HIT	2013-04-19	evaluation	1
Microsoft	Cheyenne, Laramie WY	2012	HIT	2012-05-08	permit	0
Apple	Maiden, Catawba NC	2010	HIT	2011-07-13	evaluation	1
Apple	Reno, Washoe NV	2012	HIT	2013-02-21	permit	1
Apple	Mesa, Maricopa AZ	2017	MISS	—	—	—
Amazon/Vadata	Boardman, Morrow OR	2011	MISS	—	—	—
Amazon/Vadata	Ohio, Franklin/Licking OH	2016	HIT	2017-09-01	permit	1
Yahoo	Lockport, Niagara NY	2010	MISS	—	—	—
Equinix	Ashburn, Loudoun VA	1999	HIT	2001-11-14	evaluation	2

Notes: Documented openings from the operators' published site lists; a hit is a facility under the operator's label in the same county; the gap is the EPA date's year minus the documented year, zero when inside a documented range. Source: EPA ICIS-Air and Facility Registry Service; operators' site lists; authors' calculations.

Table B2. Counties by the year of their first data center, full list

Cohort year	Counties	Cohort year	Counties	Cohort year	Counties
1983	1	2001	4	2015	4
1987	1	2002	5	2016	5
1988	1	2003	2	2017	3
1989	1	2004	6	2018	10
1991	1	2005	1	2019	4
1992	1	2006	2	2020	2
1993	1	2007	5	2021	2
1994	2	2008	5	2022	6
1995	1	2009	4	2023	4
1996	1	2011	6	2024	6
1998	2	2012	9	2025	21
1999	1	2013	9	2026	5
2000	2	2014	1	2028	1

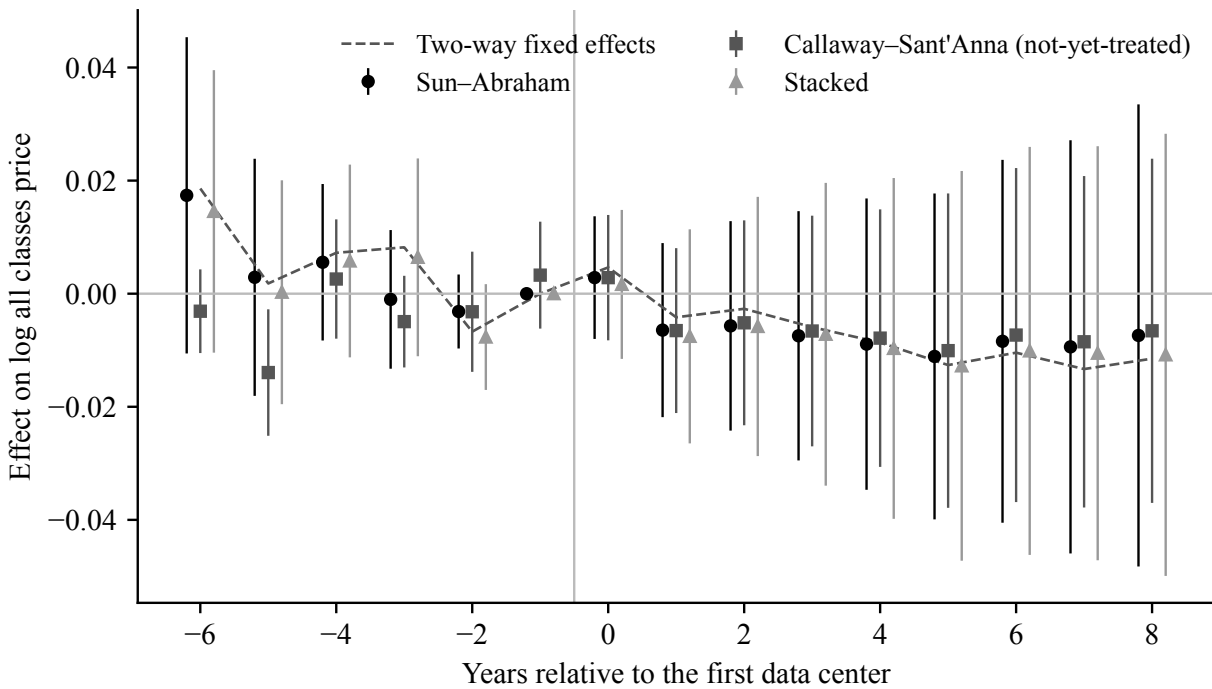
Notes: As Table 3, year by year. Source: EPA ICIS-Air; authors' calculations.

Table C1. Goodman-Bacon decomposition of the two-way fixed-effects estimate

Comparison type	Estimate	Weight
<i>Residential</i> : static two-way FE coefficient, balanced subsample	-0.0078	
earlier vs later	-0.0205	0.010
later vs earlier	-0.0019	0.009
treated vs never	-0.0077	0.981
<i>All classes</i> : static two-way FE coefficient, balanced subsample	-0.0152	
earlier vs later	-0.0210	0.010
later vs earlier	-0.0148	0.009
treated vs never	-0.0152	0.982

Notes: The static two-way fixed-effects coefficient on the balanced subsample decomposed by comparison type, with the weight on each. Source: Form EIA-861 and EPA ICIS-Air; authors' calculations.

Figure C1. Event study: the log all-classes price



Notes: As Figure 3, for the all-classes price, which includes the short-form utilities the class prices omit. Source: Form EIA-861 and EPA ICIS-Air; authors' calculations.

Table C2. Rambachan–Roth confidence sets for the residential mean effect over years 0 to 5

Restriction	95 percent confidence set (log points)
Original (years 0 to 5 average), log points	[-1.82, 1.00]
$\bar{M} = 0.5$	[-5.01, 4.55]
$\bar{M} = 1.0$	[-9.11, 8.67]
$\bar{M} = 1.5$	[-13.35, 12.95]
$\bar{M} = 2.0$	[-14.42, 14.42]
Breakdown $\bar{M}$	none

Notes: Relative-magnitudes restriction: the post-opening violation of parallel trends is at most $\bar{M}$ times the largest pre-opening violation, over the five pre-opening coefficients. Conditional least-favorable sets; no breakdown value exists because the set contains zero at every $\bar{M}$, as it must for an estimate near zero. Source: authors' calculations.

Table C3. The utility-by-state unit: the design re-run with the utility as the unit, two ways

Unit and estimator	Residential	Industrial	All classes	Pre-trend <i>p</i> (res. / ind.)	Treated / never (boot. <i>p</i> , res.)
<i>Every utility listing a treated county</i>					
Sun–Abraham, unweighted	1.03 (0.78)	6.59 (1.87)	1.07 (0.74)	0.14 / 0.08	313 / 2,080
Callaway–Sant’Anna, unweighted	1.69 (0.90)	8.62 (1.98)	1.36 (0.74)	0.23 / 0.25	313 / 2,080
Callaway–Sant’Anna, class-sales weights	0.17 (2.15)	11.01 (4.20)	3.27 (2.35)	0.11 / 0.49	313 / 2,080
Sun–Abraham, controls at least the treated 25th-percentile size	0.35 (0.66)	7.56 (2.08)	1.55 (0.76)	0.06 / 0.01	235 / 715
Median residential sales, treated / never-treated utilities (MWh)	143,882	18,669			
<i>The county’s largest utility only</i>					
Sun–Abraham, unweighted	1.75 (1.07)	8.56 (2.38)	3.31 (1.41)	0.38 / 0.00	56 / 2,080
Callaway–Sant’Anna, unweighted	2.35 (1.55)	10.28 (2.88)	3.53 (1.69)	0.60 / 0.06	56 / 2,080
Callaway–Sant’Anna, class-sales weights	0.29 (2.89)	9.24 (4.53)	2.81 (2.98)	0.67 / 0.12	56 / 2,080
Sun–Abraham, controls at least the treated 25th-percentile size	-0.39 (1.87)	8.73 (3.76)	0.95 (2.65)	0.73 / 0.07	42 / 6
Median residential sales, treated / never-treated utilities (MWh)	4,453,856	18,669			
Sun–Abraham, vertically integrated states, all never-treated utilities	2.84 (1.08)	1.93 (1.31)	—	0.07 / 0.00	41 / 1,835 (0.19)
vertically integrated states, controls at least the treated 25th-percentile size	1.12 (1.71)	0.34 (1.68)	—	0.33 / 0.04	41 / 6 (0.68)
vertically integrated states, investor-owned controls only	2.00 (1.28)	0.12 (1.51)	—	0.24 / 0.00	41 / 68 (0.90)
Sun–Abraham, restructured states, all never-treated utilities	-0.80 (1.93)	20.01 (4.99)	—	0.00 / 0.00	15 / 245 (0.81)
restructured states, controls at least the treated 25th-percentile size	—	19.96 (7.92)	—	— / 0.00	— / —
restructured states, investor-owned controls only	-0.85 (2.88)	17.48 (6.25)	—	0.00 / 0.00	15 / 22 (0.52)

Notes: Cells in log points. Unit is the utility-by-state; the outcome is the utility’s own class average revenue per kilowatt-hour (long-form filers for the class prices); the cohort is the earliest dated first opening among the counties the utility lists (territory version) or among the counties where it is the largest utility (largest version); never-treated utilities list only clean counties. Class-sales weights are the utility’s class sales in its first panel year. The pre-trend column reports the joint test for the residential and the industrial series; the bootstrap *p* in parentheses is the wild-cluster *p* on the residential cell’s static coefficient. * *p*<0.10, ** *p*<0.05, *** *p*<0.01. Source: Form EIA-861 and EPA ICIS-Air; authors’ calculations.

D Calibration details

The load share λ is computed for every treated county as $L/(Q_c + L)$ with $L = 700,800$ megawatt-hours, a 100 MW facility at an 80 percent load factor over 8,760 hours, and Q_c the state-wide total sales of the county's largest serving utility in the year before its first opening; the median Q_c is 19,571,938 megawatt-hours and the 10th and 90th percentiles are 4,175,229 and 76,917,652, over 121 treated counties with a serving utility in the sales file. The 80 percent load factor is the utilization that MIT Climate Policy Center (2025) report data centers typically run at, and that the Energy Information Administration assumed for cryptocurrency-mining facilities in its 2023 estimate, as reported in Shehabi et al. (2024).

The model's predicted change in log price for an incumbent class is $\Delta \log p = \lambda[a + g(1 - \theta\lambda)/(1 - \lambda) - \theta\kappa]$ with $a = (m - c)/p$, $g = (1 - \phi)\gamma/p$ and $\kappa = k/p$, which is equation (5) divided by the pre-existing price. The grid in Figure 9 runs a from -0.6 to 0.6 at $\theta \in \{0.5, 1, 1.5\}$ with $g = 0.10$ and $\kappa = 0.50$; the horizontal axis is the premium $a + g - \kappa$ in percent for the sales-allocation line and a for the others. The implied premium inverts $\beta = \bar{s}\lambda\pi/p$ at the median λ and $\bar{s} = 0.69$, the mean of s_τ over $\tau = 0$ to 5 at $h = 0.33$, counting the opening year as one in which a case can occur (if it cannot, as with a historical test year, the mean share is 0.54 and every inverted premium is correspondingly larger in magnitude); at $h = 0.2$ the mean share is lower and the implied premium correspondingly larger in magnitude, and at $h = 0.5$ the reverse. Under a two-channel profile in which the energy term passes through between rate cases and only the fixed-cost terms wait for one, the window-scaled tight-system prediction is 0.8 log points rather than 0.2 , and the neutral case is unchanged because its energy term is zero. The regulatory-lag shares at $h = 0.33$ are $0.33, 0.55, 0.70, 0.80, 0.86$, and 0.91 for years 0 to 5 . No moment is targeted; the calibration borrows every parameter from a cited source or computes it from the paper's own data, and the untargeted check is the sign and size of the dose-response slope against $\bar{s}\pi/p$.