Grid Lock-in and the Geography of AI*

Xinming Du† Jingyuan Guo‡

August 10, 2026

Abstract

An electricity-intensive industry that arrives first can lock in scarce local grid capacity and foreclose the next. We study this grid lock-in between cannabis cultivation and the data centers that host cloud and AI computing. Using close recreational cannabis ballots in a difference-in-discontinuities design and measuring cultivation with a new satellite detection algorithm, we find that legalization raises detected indoor cultivation by 3.7 times the sample mean within five years. Retail electricity prices rise where cultivation expands. The active data center stock falls by 0.70 facilities per county and operating capacity by 16.2 megawatts. Data centers respond by avoiding these counties, anticipating higher future prices and longer interconnection queues. The effect propagates through the shared utility grid, reaching same-grid counties even in neighboring states, which also lose data centers. Competition for shared labor is absent. A discrete-choice model of siting implies that without legalization, operators would site more capacity in the narrowly passing states, and annual electricity bills there would be 15.6 billion dollars a year lower. Slow-adjusting grids let the industry that arrives first shape where the next can build, making grid capacity a determinant of the geography of AI.

JEL Classifications: Q41, L94, L86, R11, K42
Keywords: Electricity markets, cannabis legalization, data centers, AI infrastructure, grid lock-in

*We thank David Lagakos, Chen Liu, and Wolfram Schlenker for helpful comments.
†Department of Economics, National University of Singapore. Email: xdu@nus.edu.sg.
‡Department of Economics, National University of Singapore. Email: jingyuan.guo@u.nus.edu.

1 Introduction

When a local input is scarce and slow to expand, does the industry that claims it first decide where the next one can build? Electricity is the sharpest case, because generation and transmission take years to build (Rand et al., 2024), so a fast rise in local demand raises power prices for years rather than clearing through new supply. A latecomer that needs the same power absorbs the higher cost or builds where power is cheaper (Kahn and Mansur, 2013). We study this grid lock-in in the U.S. geography of AI.

AI is the defining industry for the next decades of growth and global competition, which makes this question urgent. Data centers consume 4 to 5 percent of U.S. electricity today and could reach 9 to 17 percent by 2030 (Electric Power Research Institute, 2026). Unlike traditional heavy industry, operators are not tied to a place, building wherever land, power, cooling, and connectivity allow and reoptimizing as conditions change. A local electricity shock, including one set off by another electricity-intensive industry, can therefore move where AI capacity gets built.

Grid lock-in is hard to observe, because it is a claim about order of arrival. The industry that claims capacity first chooses where to go. When no second industry follows, we cannot tell whether it was foreclosed or was never coming. Facilities that were never built are visible only against a comparison group that would otherwise have received them. Electricity also has to be a large enough input for both industries that wages or land are less plausible reasons the latecomer stayed away.

Close recreational cannabis ballots offer one setting that meets both conditions. Indoor cultivation competes with data centers for power and for little else they need. It draws 1 percent of total U.S. electricity, and per square foot it uses eight times the energy of a typical commercial building (Mills, 2012). Where a ballot passes, that load arrives quickly, a sharply timed and localized shock to electricity demand. Close ballots make that arrival as good as random. We compare counties in states that barely passed with counties in states that barely failed near the 50 percent threshold, and the failing side supplies the counties that would otherwise have received the missing facilities.

To examine grid lock-in, we assemble a national county-year panel spanning 2000 to 2024 that links data centers, electricity markets, and cannabis ballots. Data center activity comes from the Aterio facility database, which records the location, capacity, and development dates of 5,140 facilities across 461 counties. Retail electricity prices and quantities come from Form 861 of the Energy Information Administration, and ballot outcomes from Ballotpedia, cross-checked against secretary-of-state records. The alternative-mechanism tests use employment and wages from the Quarterly Census of Employment and Wages and from County Business

Patterns. Cultivation has no comparable source, because state reporting is incomplete, inconsistent across regulatory regimes, and absent where it remains illegal. Our detector combines nightlight intensity from the Visible Infrared Imaging Radiometer Suite (VIIRS), thermal anomalies from Landsat, and rooftop indicators from Sentinel-2 into a county-year measure of cultivation activity (Henderson et al., 2012; Donaldson and Storeygard, 2016). Because we derive it independently of state reporting, it is comparable across states and over time.

We identify the effect with a panel difference-in-discontinuities, comparing counties across the majority threshold before and after the ballot (Grembi et al., 2016). The regression includes county-by-ballot and calendar-year fixed effects and a running-variable polynomial that bends at the threshold. The county-by-ballot fixed effects absorb the fixed difference between passing and failing counties, so what we estimate is how far they diverge after the ballot. The design therefore permits selection at the threshold and asks only that it be time-invariant, a fixed pass-fail gap rather than a differential trend. It also requires that a county's outcomes not depend on ballot outcomes in other states, which the cross-state grid spillovers let us examine directly. The sample runs within ten percentage points of the majority threshold and eight years of each ballot, with standard errors clustered by ballot.

In counties of states where a recreational ballot barely passed, the active data center stock falls by 0.70 facilities and operating capacity by 16.2 megawatts (MW) per county, relative to counties of states where a ballot barely failed. Passing and failing counties do not diverge measurably in the eight years before the ballot. The gap in the data center stock opens in the ballot year and holds for the eight years we observe.

Detected indoor cultivation rises by 3.7 times the sample mean within five post-ballot years, so the ballot delivers the electricity-demand shock. The increase is 0.096 square kilometers per county in the ballot year and 0.176 by the fifth, against a mean of 0.047. Every detected pixel is confirmed by the rooftop signal, so the increase is in enclosed, electricity-intensive cultivation.

Cultivation claims scarce local grid capacity, and the price of power is where that scarcity registers. A square kilometer of detected cultivation raises local electricity prices by 2.2 percent. Total electricity consumption does not move measurably, so the new load is absorbed through price rather than through new supply.

Data centers retreat not only from the counties where bills have already risen, but also from those where grid capacity is being committed. When we split counties by pre-ballot cannabis supply, prices rise only in the high-supply group, while data centers fall in both. The retreat runs ahead of the cost in time as well, with new entry falling from the ballot year before cultivation has scaled. The retreat reaches counties in other states served by the

same utility grid, where cultivation never arrives.

Labor, land, capital and shared technology could each link the two industries, and none of them carries the effect. Data centers are labor-light, and the two industries draw on different occupations, software developers on one side and farmworkers on the other. No wage or employment margin is distinguishable from zero, and the county weekly wage moves by 6.2 dollars against a pre-period mean of 745. U.S. land-use zoning leaves little room for competition over land, since it separates the industrial parcels data centers occupy from the agricultural districts that host cultivation. Capital and knowledge channels are also unlikely to be first order, since data center investment is financed nationally rather than from a local pool that cultivation could deplete, and the two industries share no common technology base. That leaves grid lock-in through the local electricity market.

A discrete-choice model of data center siting suggests an additional annual cost of 15.6 billion dollars with legalization. Each of the 2,121 facilities announced since 2015 chooses one U.S. county. Each county's chance of being chosen falls with its electricity price, so raising one county's price sends its share of new facilities to the rest. Counterfactual analysis suggests that without legalization, operators would site more capacity in the states that narrowly passed a ballot and less in Texas, Pennsylvania and Illinois, and annual electricity bills in the narrowly passing states would be 15.6 billion dollars a year lower.

This paper contributes to the literature from four perspectives. First, it identifies grid lock-in: a channel through which one industry's entry forecloses another because order of entry determines who claims scarce local grid capacity. Existing work on how entry affects incumbents centers on labor reallocation, agglomeration, and local sectoral adjustment (Greenstone et al., 2010; Moretti, 2010; Suárez Serrato and Zidar, 2016; Glaeser and Gottlieb, 2008; Allcott and Keniston, 2018; Autor et al., 2013; Black et al., 2005), or on knowledge spillovers and product market competition (Bloom et al., 2013; Moretti, 2021). Grid lock-in instead operates through the local electricity market. The industry that commits scarce grid capacity first raises the expected price and scarcity of power, foreclosing later electricity-sensitive entrants, which avoid the area rather than contest capacity already taken. This mechanism connects order of entry to carbon lock-in and the broader economics of path dependence (Unruh, 2000; David, 1985; Arthur, 1989), in which early infrastructure commitments persist because they are costly to reverse. It is quantitatively important where electricity supply adjusts slowly (Rand et al., 2024) and later entrants have locational alternatives.

Second, the paper provides rare exogenous variation in a determinant of data center location. A fast-growing body of work measures the energy and water footprint of data centers (Masanet et al., 2020; Siddik et al., 2021; Mytton, 2021) and its recent acceleration

(Shehabi et al., 2024), and policy attention has turned to their strain on local grids and ratepayers (International Energy Agency, 2024; Electric Power Research Institute, 2024). This work is largely descriptive engineering and accounting, with little causal evidence on what shapes data center siting, because the local conditions that attract data centers are correlated with unobserved local fundamentals. Our cannabis-legalization design provides a quasi-experimental instrument for one such condition, local electricity scarcity, and shows that it reallocates data center capacity across space, giving this literature an identified determinant of siting. The result also connects to the economics of automation, digital technology, and AI, where work has emphasized labor markets, productivity, and market structure (Acemoglu and Restrepo, 2018, 2020; Brynjolfsson et al., 2021; Goldfarb and Tucker, 2019) rather than the physical infrastructure on which AI runs.

Third, the results uncover a spillover of cannabis legalization onto the geography of an unrelated and economically important industry that shares the local electricity market. Existing work on the economic effects of cannabis laws, from depenalization through medical to recreational legalization, has documented effects on traffic safety and substance use (Anderson et al., 2013; Wen et al., 2015), on crime (Adda et al., 2014; Gavrilova et al., 2019), and on local housing markets (Cheng et al., 2018). The welfare accounting of legalization has to include the industries that share the grid.

Fourth, local electricity infrastructure adds a binding constraint to the location decisions of fast-growing electricity-intensive industries. Existing work on infrastructure capacity documents the role of transportation (Donaldson and Hornbeck, 2016; Donaldson, 2018; Faber, 2014) and housing supply (Hsieh and Moretti, 2019) in shaping market access, real income, and where industry locates, while rural road access mainly moves workers out of agriculture rather than expanding local production (Asher and Novosad, 2020). What binds here is the slow adjustment of generation and transmission capacity (Rand et al., 2024).

From the policy perspective, the findings show that legalization and other industry-growth policies cannot be evaluated in isolation when entrants share a slow-adjusting grid. Electricity markets do not adapt quickly to large localized demand shocks, so capacity committed to one industry raises the cost of the next entrant and can displace it altogether. Grid planning should therefore account for the opportunity cost of competing loads, and policy appraisals should include spillovers onto captive ratepayers and mobile investment. This logic extends beyond cannabis and AI to cryptocurrency mining, semiconductor fabrication, battery manufacturing, and large-scale electrolyzer production. Regions may otherwise lose emerging economic opportunities not because of weak demand or innovation, but because earlier loads have already claimed the grid.

Our paper also provides a new satellite-based data product for cannabis cultivation. It

combines nightlight intensity from VIIRS, thermal anomalies from Landsat, and rooftop indicators from Sentinel-2 into a harmonized county-year panel through 2024 (Henderson et al., 2012; Donaldson and Storeygard, 2016). Because the measure is constructed independently of state administrative reporting systems, it remains comparable across states and over time. The resulting data product can support research on the broader economic and social effects of cannabis expansion.

2 Background and Institutional Setting

2.1 Data Centers and AI Infrastructure

Data centers are the buildings, power systems, and cooling plants that house the servers behind model training, inference, and cloud services. Hyperscale campuses operated by Amazon, Google, Microsoft, Meta, and Oracle have grown from loads of tens of megawatts a decade ago to several hundred megawatts, and announced AI training campuses exceed 1 gigawatt.

Hyperscale facilities use large amounts of electricity and capital but relatively little labor. A 100-megawatt facility operating near nameplate capacity consumes 800 gigawatt-hours per year, making its annual power bill tens of millions of dollars at typical U.S. industrial rates. A campus drawing several hundred megawatts is operated by a permanent staff in the low hundreds, made up of technicians, network engineers, and security personnel, far fewer workers than a comparably sized industrial plant. Construction costs are 9.5-13.3 million dollars per megawatt in U.S. markets (Turner and Townsend, 2025), so the shell, power, and cooling of a gigawatt-scale AI campus cost 9.5-13.3 billion dollars before any computing hardware.[1] For these facilities, electricity prices and the depreciation of long-lived capital matter more for profitability than local labor costs.

U.S. data center capacity has historically concentrated in a few regional clusters. Northern Virginia, anchored by Loudoun County, is the largest, with several gigawatts of installed capacity. These clusters reflect dense fiber-optic interconnection, cheap industrial electricity, and state and local tax incentives for capital-intensive infrastructure.

Figure 1 shows the geographic distribution of U.S. data center facilities. Data centers are geographically mobile over the long run, and electricity dominates operating cost. Industrial retail rates vary by a factor of two or more across U.S. counties, and a one-cent change in the price per kilowatt-hour changes the lifetime electricity cost of a 100-megawatt facility by

[1] Announced hyperscaler capital expenditure now reaches hundreds of billions of dollars annually across the major U.S. firms.

more than 100 million dollars over a 15-year lease.

2.2 Indoor Cultivation and Electricity Demand

Indoor cannabis production relies on high-intensity lighting, cooling, dehumidification, and continuous air circulation to maintain a controlled growing environment. Mature canopies receive 12-18 hours of light per day, making electricity the largest variable operating cost. Mills (2012) estimates that indoor cultivation uses eight times as much energy per square foot as a typical commercial building. Cannabis businesses grew from 1 percent of Denver's electricity use in 2013 to 4 percent in 2018 (Summers et al., 2021). Lighting and environmental control dominate the load, placing indoor cultivation among the most energy-intensive agricultural or light-industrial activities (Summers et al., 2021).

Outdoor cultivation uses far less electricity, but regulation and product-market incentives concentrate licensed production indoors. Security, video-surveillance, and seed-to-sale tracking requirements are more easily met in enclosed facilities. Zoning steers licensed cultivation into agricultural districts and dedicated cultivation overlays, not onto the industrial parcels near substations and fiber that data centers occupy. Indoor cultivation also permits several harvests a year against a single outdoor season, tighter control of potency, and protection from climate and pests. California and Oregon retain a meaningful outdoor share, whereas licensed production in Colorado is overwhelmingly indoor. Recreational legalization expands indoor cultivation, the most power-intensive form of production.

2.3 U.S. Recreational Cannabis Legalization

U.S. cannabis policy has changed in two waves over the past three decades. Medical legalization began with California's Proposition 215 in 1996 and spread to 38 states by 2025, allowing licensed cultivation and dispensary sales restricted to patients with a physician's recommendation and a state-issued medical card for qualifying conditions. Recreational legalization began in 2012, when Colorado's Amendment 64 and Washington's Initiative 502 passed by statewide ballot, and as of 2025 has spread to 24 states and the District of Columbia. Legalization removes the medical-card gatekeeper and opens purchase and possession to any adult 21 years or older, sharply expanding the addressable market. In Colorado, recreational dispensary sales overtook medical sales within 12 months of the first stores opening in January 2014 and far exceeded medical sales within a few years, with similar trajectories in Washington, Oregon, California, and Nevada. Lawful cultivation expands well beyond the medical-only baseline, and the grid capacity it commits is the shock we study. Stores typically open 12-24 months after the ballot. Most recreational adoptions

have come through citizen ballot initiatives, a smaller number through state legislation. For the empirical design, ballot initiatives have two useful features. Their timing is determined well before election day through signature collection and certification, and passage is decided by the statewide vote rather than by contemporaneous conditions in any one county.[2]

Figure 2 maps state outcomes on the latest recreational cannabis ballot. Cannabis remains a Schedule I controlled substance under the federal Controlled Substances Act, with two implications for production geography. Interstate shipment remains illegal, so each state with recreational legalization must build out within-state cultivation capacity to supply its retail market. State regulations typically require licensed cultivators to register each plant in a track-and-trace system, maintain perimeter security, and meet building-code requirements most easily satisfied by purpose-built indoor warehouses. Recreational legalization in a given state therefore commits new within-state indoor cultivation capacity to the local grid, with the timing set by the vote rather than by the year-to-year movement of local economic conditions.

2.4 U.S. Electricity Markets and Slow Supply Adjustment

U.S. electricity is delivered through regional grids and retail markets that differ in structure. All adjust supply slowly. Wholesale generation in most of the country is coordinated through one of seven independent system operators or regional transmission organizations, which run nodal or zonal wholesale markets. Retail service comes from regulated vertically integrated utilities in most states and from competitive suppliers in the restructured states. Across both structures, adding new generation or transmission takes a long time. Natural gas plants take 3-5 years from siting to operation, nuclear units and large transmission lines take 7-15 years, and interconnection queues for new generators run to 5 years or more in the major regional markets (Rand et al., 2024). Short-run supply expansion therefore cannot absorb a localized step increase in electricity demand, which instead shows up in higher local prices, especially where transmission constraints limit imports from outside the area.

How a demand shock reaches local prices depends on market structure. Retail rates in regulated states are reset through rate cases that pass capacity, transmission, and fuel costs through to consumers with multi-year lags, so a sustained demand shock raises prices gradually. Each state's public utility commission decides those cases for service inside its own borders, so a utility whose territory crosses a state line files separate rates on each side and recovers a local cost increase only from the customers of the state that approved it.

[2]The ballot pipeline typically begins with petition drives 12-24 months before election day, with signatures verified by the secretary of state several months before the vote. Once certified, the measure appears on the next regular general election ballot.

In restructured states, wholesale locational marginal prices respond faster to transmission congestion, so a binding constraint can open sharp local price differentials within a single regional-market footprint. Even outside congested zones, capacity payments and ancillary-service costs are allocated to load-serving entities in proportion to their share of system peak, so higher local peak demand raises the per-kilowatt-hour cost borne by all local customers, including industrial users such as data centers. A localized cultivation shock thus reaches local retail prices over several years, through rate cases, congestion pricing, and capacity cost allocation.

A large new load is connected on the same slow timetable. A prospective data center applies to the utility serving the site, which studies whether the local substation, feeder, and upstream transmission can carry the requested load before it commits to serve, and the network upgrades that study triggers enter the rate base. That study and the load queues utilities publish are how a developer comparing counties learns how much local headroom is already claimed, so a county whose capacity is being committed reads as constrained years before the constraint reaches a retail rate.

Recent episodes show the constraint binding. The 2025-2026 base residual capacity auction run by PJM Interconnection, the regional transmission organization serving the Mid-Atlantic, cleared at nine times the previous auction's price. PJM attributes the increase in part to rapid load growth from data center construction in its footprint (PJM Interconnection, 2024). In eastern Loudoun County, Dominion Energy stopped connecting new data centers in 2022, and connections resumed only in part while the transmission upgrades that relieve the constraint ran to 2026. In each case the limit on new industrial load is grid capacity, not demand.

3 Data

3.1 Data Center Facilities

Facility-level data on U.S. data centers come from the Aterio data center database, a commercial industry tracker that compiles announcement, construction, activation, and withdrawal dates for each facility. Each record also carries the county location, the nameplate power capacity in megawatts, the provider name, an indicator of provider type, and an indicator of on-site generation equipment. The extract is current as of October 2025 and covers 5,140 facilities across 461 U.S. counties. The county-year panel runs 2000-2024. We use these dates to aggregate the records to the county-year level. The resulting measures capture the year-end stock at each development stage and the annual flows of entries and exits. An

active facility is one that has entered operation by year end and has not subsequently been withdrawn. We build parallel megawatt totals from nameplate capacity, imputing the median capacity in the same provider-type group when capacity is missing. The aggregation also records provider type as hyperscaler, colocation, or unknown. Provider type comes directly from the Aterio classification, and we cross-check it by hand against publicly disclosed hyperscaler portfolios for the largest operators. We compare the county-year totals with the Synergy Research Group hyperscale facility count and the Cushman & Wakefield Global Data Center Market Comparison reports at the metropolitan-area level.

3.2 Cannabis Legalization and Ballot Outcomes

We compile state recreational cannabis ballot outcomes from Ballotpedia and cross-check them against secretary-of-state records. The dataset contains thirty-four statewide recreational ballots in twenty states and the District of Columbia over 1972-2024, with the ballot year, the Yes share, the No share, the binary outcome, the ballot origin as either citizen initiative or legislative referral, and the topic.

The empirical strategy uses the Yes share minus 50 percent as the running variable, so crossing the simple-majority threshold provides the intent-to-treat assignment. For states with multiple ballots, the cross-sectional design keeps the most recent ballot, while the stacked and panel designs treat each of the thirty-four ballots as a separate event. The bandwidth of 0.10 that Section 4 holds throughout retains twenty-six ballots, twelve failing and fourteen passing. All twenty-six enter the panel design; twenty-three are observed in the first post-ballot year the pooled design uses, and thirteen enter the cross-sectional design, which keeps one ballot per state. Treatment begins in the ballot year, and licensed retail typically opens 12-24 months later. The running variable remains the ballot vote share in every post-ballot year. A county enters once for each ballot it is matched to, so the unit of observation in the stacked and panel designs is the county-ballot-year.

Two majority outcomes did not translate into retail sales. South Dakota's 2020 Amendment A received 54 percent of the vote but was overturned by a state court in 2021 before retail sales began. Florida's 2024 measure also received a majority but fell short of the 60 percent supermajority required by state law. We retain both majority outcomes in the intent-to-treat assignment, and Section 4.4 reports the estimate with South Dakota's overturned initiative counted as realized legalization.

Maryland and New Jersey legalized through legislative referrals tied to a ballot question rather than citizen initiatives, and the running variable still applies because the referral required popular approval at the same statewide threshold. Nine states legalized through

state legislation without any ballot, namely Connecticut, Delaware, Illinois, Minnesota, New Mexico, New York, Rhode Island, Vermont, and Virginia. We code them with the year of legislative passage and a missing ballot vote share. Having no running variable, they lie outside the discontinuity sample, and they enter the location-choice model of Section 6 as counties that never carried a ballot.

3.3 Cannabis Cultivation Activity

We measure indoor cannabis cultivation activity with a satellite detection algorithm that combines three remote-sensing signals at the county-year level. First, nightlight intensity from the VIIRS Day/Night Band on the Suomi NPP and NOAA-20 satellites captures the high-intensity grow lights typical of indoor cultivation. The Day/Night Band has a native spatial resolution of 750 meters at nadir and provides monthly cloud-free composites from 2012 onward via the NASA Black Marble product. Second, surface thermal anomalies come from Landsat 8 thermal infrared bands 10 and 11 at 100-meter native resolution resampled to 30 meters, with a 16-day revisit cycle and orbit-corrected brightness temperatures from the U.S. Geological Survey (USGS) Collection 2 Level-2 product. Third, rooftop indicators come from Sentinel-2 multispectral imagery at 10-meter resolution, distributed by the Copernicus program with a five-day revisit. Brightness and vegetation-index bands separate greenhouse, warehouse, and industrial-cultivation rooflines from residential and conventional commercial roofs, and Sentinel-2 coverage sets 2017 as the first year of the three-signal measure.

We combine the three signals using inverse-variance weights estimated on state-licensed cultivation facilities in Colorado, Oregon, and Washington. The held-out composite achieves an area under the receiver operating characteristic curve of 0.87, compared with 0.76-0.82 for the individual signals. The primary continuous outcome is the area, in square kilometers, of pixels where all three signals fire in the same county-year, formed by summing flagged pixel area and converting from square meters. We also report the composite activity index. Two binary classifiers built from the same detections split counties in the mechanism analysis, an any-detection classifier that marks a county-year with any triconfirmed pixel and a mean-cut classifier that marks a county-year whose two-signal detected area exceeds the sample mean.

Apart from the licensed-facility validation, the measure does not depend on state administrative reporting, so the same detection rule applies in every county and year, including the states where cultivation is never licensed. The three-signal panel covers 2017-2024, since the rooftop signal begins in 2017. The two-signal nightlight-thermal measure covers 2000-2024 and carries the earlier ballots, taking DMSP-OLS nightlight with MODIS land surface temperature before 2012 and VIIRS with Landsat thermal from 2012 onward, and

calendar-year effects absorb the sensor transition, with all annual analyses ending in 2024. Appendix A documents each signal's processing pipeline, the licensed-facility validation, and the geographic distribution of detected cultivation.

3.4 Electricity Prices and Consumption

Retail electricity prices and quantities come from the U.S. Energy Information Administration (EIA) Form 861 utility-by-state annual data. Form 861 reports retail revenues and sales by customer class, covering residential, commercial, and industrial service. We compute each class price as that class's own revenue divided by its own sales, in nominal dollars per kilowatt-hour, so a shift in the mix of load across residential, commercial, and industrial service cannot move the class prices we report, and calendar-year fixed effects absorb inflation common to all counties.

We map utilities to counties with the service-territory schedule of Form 861, which lists the counties each utility serves in each state and year, and we match those county names to Census FIPS codes, so the assignment is rebuilt for every year of the panel. Franchised retail service territories are set by state commissions and change rarely, so the mapping does not move with the outcomes. Service territories routinely cross county and state lines. Form 861 reports revenue and sales by utility and state, so we allocate each utility's state total across the counties it serves that year in proportion to their 2000 Census population, a benchmark fixed before every ballot the electricity panel covers. We then sum allocated revenue and sales across all utilities serving a county, and the ratio of the two defines its price.

Local consumption is the same allocated Form 861 panel, aggregated to county-year megawatt-hours overall and by customer class. Because sales are reported by utility and state, the quantity variation a county inherits is that of the service territories serving it. We also build a county-to-balancing-authority crosswalk from the balancing-authority schedule of Form 861, which defines the wider grid footprint used in Section 5.4. The full electricity panel covers 2000-2024.

3.5 Labor Markets

Labor-market outcomes used in the alternative-mechanism tests come from two complementary sources, the Bureau of Labor Statistics (BLS) and the U.S. Census Bureau. The first is the BLS Quarterly Census of Employment and Wages (QCEW), which reports covered employment and wages by county, industry, and ownership, compiled from establishment unemployment-insurance records. We obtain the QCEW at the six-digit North American Industry Classification System (NAICS) level and aggregate it to the county-year level for

total employment and two-digit sectors. We report the broad sectors of the local economy separately, along with the data center, cannabis cultivation, and shared building-trade industries themselves, taken as NAICS 518210, NAICS 111419, and NAICS 2382. A county-year with no covered firm in an industry enters as zero, as does a cell withheld for disclosure, so the narrow industry columns respond to the presence of a covered firm as well as to its payroll and we read them beside the county aggregate, which is complete in every county-year and carries the falsification. QCEW covers 95 percent of U.S. wage and salary employment.

The second is County Business Patterns (CBP), an annual county-by-industry tabulation produced by the U.S. Census Bureau from the Business Register and available from 1986 onward. CBP reports the number of establishments and mid-March employment in each six-digit NAICS industry by county and by establishment-size class. The CBP employment and establishment counts track the QCEW counts at the NAICS aggregations we use, drawing on the imputed CBP panel that fills the county-industry cells the Census Bureau withholds for disclosure. Occupational composition comes from the BLS Occupational Employment and Wage Statistics, using the May 2024 national release by industry.

4 Empirical Strategy and Baseline Effects

A naive comparison of data center activity between legalizing and other places does not recover the causal effect of legalization. States legalize when local economic conditions, political climate, and industry composition move in ways that also shape where data centers locate. This makes the legalization indicator endogenous to the very forces that drive siting, and a simple regression confounds the two. We break this endogeneity with close statewide ballot margins on recreational legalization. Whether a ballot barely passes or barely fails is plausibly unrelated to the local economic conditions that drive data center siting, yet it sharply determines whether the state legalizes recreational cannabis cultivation. The discontinuity at the passage threshold therefore isolates the causal effect of legalization. Let s index states. For a state's ballot we define the running variable as the Yes share minus the 50 percent passage threshold

$$R_s = \mathit{YesShare}_s - 0.50 \tag{1}$$

and passage as $\mathit{Pass}_s = \mathbf{1}\{R_s \geq 0\}$. We build the design up across three stages that progressively tighten identification, from a single-ballot cross-section to a dynamic panel. Each stage addresses what the previous one leaves open, and Section 5 reuses the design throughout the mechanism analysis. Throughout we hold the bandwidth at $|R_s| \leq 0.10$ around each cutoff and cluster by ballot.

4.1 A Single-Ballot Cross-Section

The first stage is a cross-sectional regression discontinuity (RD) using outcomes measured one year after each state's most recent ballot,

$$Y_{cs} = \beta\, Pass_s + \gamma_1 R_s + \gamma_2 (R_s \times Pass_s) + \varepsilon_{cs} \tag{2}$$

where c indexes counties. The running variable enters as a linear polynomial that bends on each side of the cutoff, and the model has no fixed effects. The coefficient β is then the level contrast in active data center stock between barely-passing and barely-failing counties. States that never held a recreational cannabis ballot have no running variable and drop out of the sample.

The single-ballot column of Appendix Table C1 and the full outcome set in Appendix Table C2 report these estimates. Every outcome points the same way. Counties in barely-passing states host 1.36 fewer active data centers per county than counties in barely-failing states, install 20 fewer megawatts of operating capacity, take in fewer new entries, and host fewer facilities with on-site generation. Restricting each state to its single most recent ballot leaves only thirteen ballots in the estimation sample, too few to estimate the pass-fail contrast precisely.

4.2 Pooling Across Ballots

To increase the number of ballots near the cutoff, in the spirit of the multiple-cutoff pooling of Cattaneo et al. (2020b), the second stage stacks every ballot event, indexed by a, and estimates the pooled cross-section,

$$Y_{csa} = \beta\, Pass_{sa} + \gamma_1 R_{sa} + \gamma_2 (R_{sa} \times Pass_{sa}) + \varepsilon_{csa} \tag{3}$$

still without fixed effects. Each ballot keeps its own re-centered running variable, so continuity is required only locally at each threshold, and the pooled coefficient is a weighted average of ballot-specific intent-to-treat contrasts.

Pooling brings twenty-three ballots into the cross-section and sharpens every estimate.[3]

[3]The bandwidth contains twenty-six ballots, twelve failing and fourteen passing. Because this cross-section is measured one year after each ballot, as in Section 4.1, and the panel ends in 2024, the three ballots decided in November 2024, in Florida, North Dakota, and South Dakota, have no 2025 observation to enter it, leaving twenty-three. The panel design of Section 4.3, which uses every post-ballot year, retains all twenty-six. With only two dozen ballot clusters even after pooling, conventional cluster-robust standard errors can over-reject. We therefore do not rely on them alone. Section 4.4 confirms the headline estimate with the wild cluster bootstrap, which retains near-nominal size at this number of clusters.

Appendix Table C3 reports the pooled estimates. Counties in barely-passing states host 1.46 fewer active data centers per county, install 19 fewer megawatts of capacity, take in fewer new entries, and host fewer generation-equipped facilities, all now precisely estimated. The active-stock contrast is larger, and every contrast far more precise than its single-ballot version, confirming that the cross-sectional gap is a feature of the full sample rather than one ballot. Whether passage opened the gap or the gap came first is what the panel design settles.

The pooled estimate identifies a causal effect only if there is no selection into treatment or control. Barely-passing and barely-failing states must be comparable in data center activity apart from the ballot outcome itself, which the close margin makes plausible. We examine this by estimating equation (3) separately at each event time, where a pass-fail gap already open before the ballot reveals a pre-existing difference rather than an effect of passage.

Figure 3 shows the assumption fails. The contrast is already open before the ballot, with passing states hosting between 0.8 and 1.0 fewer facilities across the pre-ballot years, and it turns sharply more negative after passage, reaching 1.5 fewer facilities by the first post-ballot year and 4.5 fewer by the eighth as the effect accumulates. Because the cross-sectional design blends this pre-existing difference with the effect of passage, it overstates the causal effect. The violation nonetheless takes a specific form. The pre-ballot gap is a stable level difference rather than a diverging trend, holding with a nearly flat slope, and a fixed difference is exactly what differencing removes. This observed selection is why we move to a panel difference-in-discontinuities (Grembi et al., 2016).

4.3 The Panel Difference-in-Discontinuities

The third stage absorbs the fixed pass-fail gap by combining the discontinuity with a difference-in-differences (DiD) panel

$$\begin{aligned} Y_{csta} = \beta\left(Pass_{sa} \times Post_{sta}\right) + \gamma_1 Post_{sta} + \gamma_2\left(R_{sa} \times Post_{sta}\right) \\ + \gamma_3\left(R_{sa} \times Pass_{sa} \times Post_{sta}\right) + \alpha_{ca} + \lambda_t + \varepsilon_{csta} \end{aligned} \tag{4}$$

where α_{ca} are county-by-ballot fixed effects, λ_t are fixed effects for calendar year t, and $Post_{sta} = \mathbf{1}\{t \geq B_{sa}\}$ indicates the years at or after ballot year B_{sa}. The county-by-ballot fixed effects absorb every time-invariant county characteristic specific to a ballot's comparison group, so β identifies the post-ballot change in the pass-fail contrast rather than its level. We keep observations within eight years of each ballot, bounding the event horizon around

each vote (Cengiz et al., 2019).[4]

Identification now rests on a parallel-trends assumption, conditional on the running-variable polynomial. The design permits selection at the threshold and asks only that it be time-invariant, a fixed pass-fail gap and not a differential trend. Section 4.2 documented selection of that kind. It also requires the stable unit treatment value assumption (SUTVA) that a county's outcomes do not depend on ballot outcomes in other states. We examine that assumption directly through the cross-state grid spillovers of Section 5.4.

To test this parallel-trends assumption and date the onset of the effect, the dynamic form of the same panel replaces the single post-ballot term with a full set of event-time interactions, with event time defined as $\tau_{sta} = t - B_{sa}$

$$\begin{aligned} Y_{csta} = & \sum_{k \neq -1} \beta_k \left(Pass_{sa} \times \mathbf{1}\{\tau_{sta} = k\} \right) + \gamma_1 \, Post_{sta} + \gamma_2 \left(R_{sa} \times Post_{sta} \right) \\ & + \gamma_3 \left(R_{sa} \times Pass_{sa} \times Post_{sta} \right) + \alpha_{ca} + \lambda_t + \varepsilon_{csta} \end{aligned} \tag{5}$$

with the year before the ballot as the omitted reference, so the coefficients β_k trace the effect of passage at each lead and lag. The pre-ballot leads are the direct test of parallel trends, since under a valid design the β_k for $k < -1$ are jointly indistinguishable from zero and the effect emerges only after passage.

Figure 4 shows both. The pre-ballot coefficients are small and jointly indistinguishable from zero, so the parallel-trends test is not rejected. The active data center stock then falls in the ballot year by 0.64 facilities and stays 0.53-0.69 facilities lower at every horizon through the eighth post-ballot year. The baseline crowd-out thus emerges within the year of passage and persists without fading. Forward-looking siting produces this pattern.

In the static specification, the active data center stock falls by 0.702 facilities per county (Appendix Table C1).[5] The same decline appears on every other measure of data center presence in Appendix Table C4. Operating capacity falls by 16.2 MW, entrant capacity by 4.4 MW, and generation-equipped facilities by 0.445, each precisely estimated, and new entries fall by 0.089 per year in the same direction. At a construction cost of 10-20 million dollars per megawatt (Section 2.1), the lost 16.2 MW is 160-320 million dollars of investment a legalizing county does not receive. This forgone investment is not an unambiguous welfare

[4]Each stacked ballot draws its controls from a single state, so $Pass_{sa} \times Post_{sta}$ varies at the year-by-ballot level and would be absorbed by year-by-ballot fixed effects. We therefore use calendar-year fixed effects alongside the county-by-ballot fixed effects, which restores identification of β.

[5]The 0.702-facility loss exceeds the mean stock of 0.446 across bandwidth counties because the effect is identified off the counties that host data centers, not the many that never do. A tenth of bandwidth counties host an active data center at some point, and among those the average stock is four or five facilities, well above the estimated loss.

loss, since data centers bring few local jobs and add to the local electricity load, but the county still forgoes real economic activity and the property-tax base a data center anchors.

4.4 Robustness

The close-election design requires that no state can place its vote share precisely on either side of the fifty-percent threshold, so that barely-passing and barely-failing states are comparable at the margin (Lee, 2008). Such precise control over the vote share is implausible for a statewide ballot, whose margin aggregates millions of independent votes that no campaign or interest group can steer finely enough to place the state on a chosen side of the cutoff. Sorting need not be deliberate, however. Caughey and Sekhon (2011) find that bare winners and losers of close U.S. House races differ on pre-determined characteristics even absent manipulation, though such imbalance is the exception rather than the rule across the more than forty thousand close elections examined by Eggers et al. (2015). Our design does not require balance in levels. Section 4.2 finds a pass-fail gap already open before the ballot, and the panel of Section 4.3 differences it out.

We examine this assumption directly. Sorting across the threshold would leave a discontinuity in the density of the running variable at the cutoff, with too many states landing just on one side. The local-polynomial density test of Cattaneo et al. (2020a), a refinement of McCrary (2008), finds no such discontinuity on the twenty-six ballots within the bandwidth and on all thirty-four.[6] Appendix Figure B1 plots the estimated density with its confidence bands. The running variable passes smoothly through the cutoff, consistent with as-good-as-random assignment of passage.

A cannabis ballot may also signal voter willingness to adopt other regulation that independently affects data center siting. We test for such policy bundling by re-estimating the panel difference-in-discontinuities of Section 4.3 with five other state policies as outcomes, namely a binding mandatory Renewable Portfolio Standard (DSIRE), Regional Greenhouse Gas Initiative membership (RGGI), a right-to-work statute (NCSL), full retail electricity choice (EIA), and the top statutory corporate income tax rate (ITEP).[7] None jumps at the cutoff after passage (Appendix Table C5). Only RGGI membership moves, by 6 percentage points, and only at the 10 percent level. The right-to-work and renewable-portfolio coefficients are within two hundredths of zero. The crowd-out is therefore not a bundled-policy

[6] The test statistic is -0.52 ($p = 0.60$) and -0.75 ($p = 0.46$), respectively.

[7] We select these as the state policies that could accompany a cannabis ballot while also shifting data center location through its main input costs of electricity, labor, and taxes. The renewable-portfolio standard, RGGI carbon pricing, and retail-choice deregulation bear on electricity prices, right-to-work status on labor cost, and the corporate income tax rate on the tax burden.

effect. No state in the bandwidth changes its retail-choice status during the window, so that coefficient is a mechanical zero, and the corporate rate moves by less than a percentage point, a change statistically indistinguishable from zero.

A pre-trend test that fails to reject is not by itself evidence that trends are parallel (Rambachan and Roth, 2023). The pre-ballot coefficients in Figure 4 are individually small but drift upward at the farthest leads, so we ask whether a lingering difference in trends, rather than passage, could account for the post-ballot decline. Even allowing a difference in trends as large as the largest one seen across the seven pre-ballot years, the robust confidence set for the average post-ballot effect still excludes zero, and admits zero only for larger differences (Figure 5). The crowd-out is therefore not an artifact of differential pre-trends.

The control group is itself partly treated because three control states legalize within their own comparison windows: California in 2016 after its 2010 ballot failed, Oregon in 2014 after its 2012 ballot failed, and Arizona in 2020 after its 2016 ballot failed. Dropping each control state's observations from the year it legalizes onward, in the spirit of the not-yet-treated comparisons of Callaway and Sant'Anna (2021), leaves the crowd-out at 0.46 fewer facilities, and at 0.56 when South Dakota's overturned 2020 initiative is also counted as legalization (Appendix Table C6). The estimate survives, and the smaller magnitude reflects composition rather than contamination. Contamination biases the pass-fail gap toward zero, so removing it should, if anything, enlarge the estimate. Instead the excluded states supply three of the six control ballots at the seventh and eighth post-ballot years, exactly where the accumulated gap is widest. At a three-year window the restricted and baseline samples give nearly identical estimates, 0.192 against 0.196, and they diverge only as the window extends into those late years.

The magnitude might be an artifact of discretionary specification choices, namely the length of the event window, the width of the bandwidth, and the fixed effects. The level effect on the active stock grows monotonically with the event window, from 0.196 fewer facilities per county at a three-year window to 0.496 at six years and 0.702 at eight, and remains precisely estimated at every window (Appendix Table C7). The growth is the accumulation of a permanent effect rather than sensitivity to tuning. Wider windows take in the later post-ballot years where the pass-fail gap is widest, while the dynamic pattern is unchanged, with flat pre-ballot coefficients, a sharp break at the ballot year, and a stable post-ballot path. The estimate is likewise stable across bandwidths, staying between 0.64 and 0.84 fewer facilities over the 0.07-0.15 range and tightly estimated throughout, and it grows to 1.19 fewer facilities at the mean-squared-error optimal bandwidth of 0.04, where the discontinuity is identified most sharply (Appendix Table C8).[8] It is also stable across

[8]The optimal bandwidth of 0.04 leaves only four or five ballots on each side of the cutoff, too few for

fixed-effect structures, spanning 0.68-0.72 fewer facilities across county, county-by-ballot, and county-by-ballot-plus-year fixed effects once the ballot-specific post indicator is included (Appendix Table C9). No single specification choice drives the result.

Because the design fits a straight line on each side of the cutoff, a curved relationship between vote share and data center activity could imitate a jump. Adding a quadratic term on each side moves the estimate to 0.99 fewer facilities, against 0.702 at baseline (Appendix Table C10). The crowd-out is not an artifact of the linear fit, and the quadratic makes it larger, not smaller. We go no higher than a quadratic, since high-order polynomials are unreliable in this design (Gelman and Imbens, 2019).

A final concern is the reliability of the standard errors. The design compares only a dozen ballots on each side of the cutoff, and with so few clusters the conventional cluster-robust standard errors can over-reject (Cameron and Miller, 2015). We therefore confirm the headline estimate with the wild cluster bootstrap, which stays reliable when clusters are few (Cameron et al., 2008), reported in Appendix Table C11. The active-stock effect remains robust under the bootstrap, so the crowd-out does not rest on the few-cluster approximation.

5 Mechanism

5.1 The Cannabis Cultivation First Stage

This section examines the channels through which legalization could operate, beginning with the demand shock that cultivation places on the local electricity market. We apply the dynamic panel design of equation (5) to satellite-detected indoor cultivation area, the three-signal measure of Section 3.3. Because it is built independently of state administrative reporting, it is comparable across states with different regulatory regimes.[9] The measure is sparse. Detected cultivation appears in a small fraction of county-years, and a typical county-year is at zero, so the event study reads the post-ballot shift in the distribution rather than a change at the median.

Figure 6 shows the first stage. The pre-ballot coefficients are small and jointly indistinguishable from zero, so the falsification test gives no evidence against parallel trends. Detected cultivation rises steadily from the ballot year onward, with the per-year effect climbing from 0.096 square kilometers at passage to 0.176 square kilometers by the fifth

reliable inference, so we hold the baseline at 0.10, which keeps a dozen.

[9]Only legalizing states keep licensing records, so administrative data cannot measure cultivation in the failing states that form the control group. A post-ballot rise in detected area could also reflect a detector that sharpens over time, but calendar-year fixed effects absorb any change in detector performance common to all counties, and the estimate compares passing to failing states within the same year.

year, significant at the 5 percent level. Against a sample mean of 0.047 square kilometers, the post-ballot increase is 3.7 times the average detected area. Appendix Table C12 reports the static three-stage summary behind the figure. Because our measure requires nightlight, thermal, and rooftop confirmation together, the expansion it detects is indoor cultivation. Indoor grow operations use eight times the energy per square foot of a typical commercial building, as Section 2.2 documents. Even the modest detected footprint is therefore a concentrated local load.

5.2 The Electricity-Market Channel

The cultivation shock transmits to data centers through the local electricity market. We measure the transmission in two steps. First we relate retail electricity prices to detected cultivation within county, with county-by-ballot and calendar-year fixed effects. The within-county movement of detection identifies the cultivation-price link net of permanent county differences and common year shocks. Then we estimate the panel design of equation (4) with a cultivation-intensity interaction. The interaction lets the post-ballot price effect vary with the square kilometers of detected cannabis in a legalized county, so we can ask whether prices rise most where cultivation expands most. Ballot passage identifies the average effect, and the interaction reads the price response against realized cultivation intensity; the within-county correlation shows the same pattern.[10]

Within county, retail electricity prices rise with detected cultivation. In Table 1, a one-unit increase in the inverse hyperbolic sine of detected cultivation area is associated with approximately 5.1 percent higher average retail prices in Column (1), 4.6 percent higher commercial prices in Column (2), and 7.0 percent higher industrial prices in Column (3). Commercial and industrial service are the two customer classes from which a data center buys electricity, and the association appears in both, consistent with indoor grow operations competing for the same local grid capacity.

Table 2 confirms the same pattern in the ballot design. In Panel A the average post-ballot price effect is positive but not distinguishable from zero. The cultivation-intensity interaction is positive in every column, at 2.2 percent per square kilometer of detected cannabis on average and 2.3 percent for commercial customers, both significant at the 1 percent level,

[10]Detected cultivation is itself a post-ballot outcome, so the interaction measures the price response at realized cultivation intensity rather than at an exogenously assigned dose. The within-county correlation and the cultivation-intensity interaction use the 2017-2024 period over which the three-signal satellite measure is available. Restricting to those years leaves 19 of the 26 ballots in the estimation sample at the five-year window and 23 at the eight-year window, so the estimates on the restricted sample are clustered on fewer ballots than the data-center results of Section 4.4. The ballot effects without the interaction are estimated on the full sample as well.

and 2.6 percent for industrial customers, significant at the 10 percent level. Legalized counties with more cultivation thus face larger post-ballot price increases. That locates the channel in local supply constraints rather than statewide tariff changes, since a statewide tariff change would move prices in every county of a legalizing state whether or not cultivation appears there. The ballot average is small for a mechanical reason. Detected cultivation appears in only 4 percent of county-years and remains a small share of county demand, so averaging over legalized counties averages mostly over counties with no cultivation in them.

Prices move with cultivation and aggregate quantities do not, the pattern we expect from a slow-adjusting supply side. Panel B shows that the average post-ballot effect on consumption is indistinguishable from zero. Commercial consumption is the exception, rising 3.7 percent per square kilometer of detected cannabis and significant at the 5 percent level, which is the margin on which licensed grow operations themselves buy power. Figure 7 reproduces the quantity null in the dynamic design, where pre-trends are flat and the post-ballot path drifts modestly positive without ever becoming distinguishable from zero. Slow-adjusting generation cannot relieve a local supply constraint within our horizon (Section 2.4), so the pressure shows up in price. Cultivation is large against the marginal grid capacity that prices and siting decisions turn on, and small against what a county consumes in total.

Three specification choices sit behind the price and quantity results: the detection sample, the functional form of cultivation, and the event window. The detection measure begins in 2017, so the within-county cultivation and cultivation-intensity estimates use 2017–2024 county-years, while the non-interacted ballot-price estimate also uses the full panel. The full-panel price estimate is nearly unchanged (Appendix Table C13). The functional form is not driving the within-county correlation. It holds under either transform (Appendix Table C14). Widening the event window to eight years leaves the within-county correlation unchanged (Appendix Table C15). The cultivation-intensity interaction is about half as large, and the consumption null holds in both the static and dynamic designs (Appendix Table C16 and Appendix Figure B2). The electricity channel is therefore a cultivation effect that reaches data centers through price.

5.3 Expected Resource Competition

Data center siting is decided on horizons of a decade or more, so operators respond not only to the price they observe today but to the resource constraint they expect legalization to create. Contemporaneous resource competition predicts that data centers leave where power has actually become more expensive, so the retreat should stop where the price stops moving. Expected resource competition predicts that they leave where they expect the constraint to

tighten, so the retreat should follow the policy or the grid, reaching counties whose own prices do not move.

We separate this anticipatory channel from contemporaneous cost competition by splitting both the price effect and the data center effect on the scale of cannabis supply already present before the ballot. An established supply base reveals a market and makes it cheap to serve legally, since growers, buildings, and distribution are already in place. The contemporaneous energy cost that post-ballot cultivation drives should therefore concentrate where pre-ballot supply was large. We split counties by pre-ballot cultivation or by pre-ballot cannabis manufacture and sale.

The crowd-out concentrates where cannabis supply was already established, and the price increase points the same way. Figure 8 splits on satellite-detected cultivation before the ballot, and Figure 9 on pre-ballot arrests for cannabis sale and manufacture.[11] The post-ballot price increase is positive in the high-supply group and flat in the low-supply group on both classifiers, so it is cultivation rather than legalization that raises local electricity prices. The data-center retreat is markedly larger where pre-ballot supply was higher, yet the low-supply counties also lose data centers even though their price does not move. This is what an anticipatory response predicts and a contemporaneous cost response does not. Developers price in higher future electricity costs and deeper interconnection queues as cultivation commits local grid capacity, so entry falls before any price moves on the ballot day. The split holds under both a single pre-ballot year and an average over all pre-ballot years, and a per-capita version of the splitter gives the same direction, so it is not an artifact of the detection window or of county size.

5.4 Grid Transmission and Spillovers

To locate where the price effect travels, we widen the unit from the county to the electricity grid. The grid is also where we can test SUTVA (Section 4.3), since a shared grid is how one state's ballot would reach another state's control counties. For each ballot we take every county served by a utility that also serves the balloting state, and partition this grid footprint into counties that legalize and cultivate, counties that legalize but do not cultivate, and counties in other states on the same grid that cannot cultivate, whose composition is summarized in Table 3. Table 4 reports the price results and Table 5 the data-center results.

[11] The arrest measure reaches back to the 1970s and so classifies all 26 ballots, including the two 2012 legalizers that satellite detection cannot see before 2017. Arrest counts reflect enforcement intensity as well as underlying activity, so a county that polices cannabis supply more aggressively records more arrests at the same level of cultivation. We use the measure to split the sample rather than as an instrument, since it would not satisfy an exclusion restriction.

The price effect is a within-state grid phenomenon. In the continuous design, an additional square kilometer of cultivation in other same-state counties on the grid raises a county's own retail price by 0.9 percent, while the same square kilometer of cultivation in other-state grid neighbors moves prices by 0.14 percent, an effect indistinguishable from zero. The design could have detected a cross-state spillover 70 percent as large as the within-state estimate, so the cross-state null is a quantitative bound rather than the limit of what the design can detect, and cross-border leakage does not bias the within-state estimate. Table 6 reports the within-state and cross-state estimates with their confidence intervals and minimum detectable effects on the utility and balancing-authority grids. The utility grid is the sharper footprint. On the wider balancing-authority grid both the within-state and cross-state doses collapse to economically negligible magnitudes, because the balancing authority averages over many distant counties that share no local grid constraint, so we read the utility estimate as the primary measure. Retail rates are set by state public utility commissions rather than at the wholesale level (Borenstein and Bushnell, 2015), so a cultivation shock raises retail prices within the legalizing state's own regulatory jurisdiction but not across the state line, even on a shared grid.

The data-center crowd-out is grid-wide, and it crosses the state line where the price does not. Table 5 reports a fall of 0.30 facilities in legalizing counties that cultivate, 0.35 in legalizing counties that do not, and 0.42 in other-state counties on the same grid, with the two non-cultivating groups the more sharply estimated. The within-state group sees both prices rise and data centers leave, so its crowd-out could combine a contemporaneous cost channel with the forward-looking response, and this design does not separate the two. The other-state grid group sees no price movement yet loses data centers by more, which cannot be a price response and is instead the forward-looking retreat, with developers avoiding the shared grid because they expect it to tighten. Interconnection queues and the grid headroom behind them cross the state line (Rand et al., 2024). The retail price does not.[12] The cross-state grid result therefore confirms that reading with independent variation, and the within-state result is consistent with it.

The spillover results indicate that the violation of SUTVA leads to underestimation for data centers. On the one hand, a data center may be displaced from a legalizing state to land in control counties, leading to overestimation. On the other hand, on a shared grid, the anticipatory retreat reaches the control county as well, creating underestimation. Our estimates suggest that the latter dominates the former, at least among counties that share

[12] These counties carry the same county-by-ballot and calendar-year fixed effects as the rest of the panel, so the coefficient measures how their stock changes after a neighboring state's ballot, not a standing level difference. A shock common to the whole grid would have moved their prices too.

a grid.

5.5 Ruling Out Labor Competition

The first alternative to grid lock-in is labor competition. An expanding cultivation sector would raise local wages, and data centers facing a higher wage bill would locate elsewhere. That route reaches the same crowd-out with no electricity capacity committed, which is why it is a potential explanation to be examined.

Labor competition is mechanically implausible, since neither industry is large enough to move a county wage and the two draw on different occupations. Appendix Table C17 reports the industry scale. Indoor cannabis cultivation employs 2.6 workers per county-year, 0.003 percent of county employment, and data centers employ 110.9 workers, 0.042 percent, neither approaching the labor-intensive benchmarks of manufacturing at 11.3 percent, retail at 11.4 percent, accommodation and food at 7.0 percent, or construction at 4.6 percent. The two industries also draw on separate occupational pools. Appendix Table C18 shows data center employment concentrated in computer and business work, with software developers alone at 16.1 percent, and information-systems managers, support specialists, and systems analysts also high on the list. The cultivation-side pool is concentrated in manual farm work, with farmworkers at 55.4 percent followed by hand packagers and agricultural equipment operators. Only generic office-clerk and manager roles appear among the top occupations of both, so the two industries do not bid for the same workers in operations.

Construction is the one phase where the two industries might overlap. Electrical and HVAC contractors equip both, a cultivation facility with lighting circuits and climate control and a data center with power distribution and cooling. The overlap is thin even so, because Waugh (2026) documents that the AI build-out runs on imported electrical-power and cooling equipment, leaving the installation crew as the only local labor at stake.

The falsification confirms the descriptive picture. In Table 7 no wage column rises. The county-total weekly wage moves by 6.2 dollars against a pre-period mean of 745 dollars, under one percent of the mean, the data center wage by 195.6 dollars, the cannabis cultivation wage by 70.7 dollars, and the HVAC and electrical contractor wage by 76.9 dollars, none distinguishable from zero and the sector estimates mildly negative rather than positive.[13] The same null holds across the broad sectors of the local economy. Appendix Table C19 reports the weekly-wage effect in construction, manufacturing, retail, information, accommodation and food, leisure, and professional services, and no sector shows the positive response la-

[13]The data center, cannabis, and contractor columns are filled with zero for county-years that have no operating firm, so these columns test firm entry as much as wage competition, with the intensive-margin estimates conditioning on operating firms discussed below.

bor competition requires, with the few precisely estimated coefficients negative rather than positive. Employment is equally unmoved, with the county total, data centers, cannabis cultivation, and those contractors all indistinguishable from zero, so no additional workers are pulled into any of them. Figure 10 traces total county employment in the dynamic design and shows a flat path with no break at the ballot, the precise null around legalization that the falsification predicts. The labor-competition channel requires a positive response on at least one of these margins and produces none.

The one wage that does move post-ballot is the data center wage, and it moves down rather than up. Figure 11 traces flat pre-ballot coefficients giving way to a post-ballot decline of 220 dollars a week before attenuating, with no positive response at any horizon. The two channels predict opposite signs here. Labor competition would force data centers to raise wages to retain workers, whereas a demand contraction, in which grid lock-in forecloses new data center entry, lowers the sector's demand for labor and pushes its wage down. We observe the decline, not the rise that labor competition would require, so the negative coefficient is the demand-side footprint of the same crowd-out that produces the 0.702 loss in the active stock, not evidence of competition for labor. The cannabis side shows the mirror image. Conditioning on county-years with an operating cultivator, the indoor cultivation wage rises by 619 dollars a week after legalization, in Appendix Table C20, the own-sector demand increase that legalization produces as it expands cultivation. Both wage movements are own-sector demand shifts, the data center sector contracting and the cultivation sector expanding, and neither is the labor competition that would bid wages up across the shared local market.

5.6 Other Potential Channels

Three further channels could each produce the same crowd-out without electricity: competition for land, competition for capital, and knowledge spillovers. None does.

Land is the first. U.S. land-use institutions leave little room for competition even before any estimate. Data centers and indoor cultivation occupy different rungs of the local land-use hierarchy. Data centers site on industrial and commercial parcels chosen for proximity to high-voltage substations and fiber backbone, usually inside industrial parks, where land is a small share of a capital outlay dominated by power and equipment. Licensed indoor cultivation is instead steered by state and municipal zoning into agricultural districts and designated cultivation overlays, under setback, security, and building-code requirements that purpose-built rural and light-industrial warehouses satisfy most cheaply, as Section 2.2 describes. The two uses fall under separate zoning categories and rarely bid for the same

parcels, and the cultivation footprint is in any case far too small to move a county land market.

Capital is the second, and data center investment is not raised locally. The facilities are financed by national and global hyperscalers drawing on national capital markets, as Section 2.1 describes, so the physical presence of cannabis cultivation in a county does not compete for the funds that build data centers. The hyperscaler capital expenditure that builds these facilities runs to hundreds of billions of dollars a year and is allocated across the country by a handful of firms, a pool that a single county's cannabis sector cannot deplete. That capital is mobile and chooses where to locate, the margin the siting model of Section 6 captures, not a local fund that cultivation could crowd out.

Knowledge spillovers are the third, and they cannot bind because cannabis cultivation and data centers share no common technology or knowledge base through which one could raise or lower the other's productivity (Bloom et al., 2013; Moretti, 2021). Indoor growing is a horticultural and climate-control activity while data centers are computing infrastructure, so neither produces the know-how the other uses, and legalization cannot reshape data center siting through a knowledge channel.

The broad-sector falsification of Section 5.5 shows that the crowd-out is not a general consequence of legalization for the local business climate. No local sector, from construction and manufacturing to retail, information, and professional services, shows a positive wage or employment response to legalization, so the result is not a broad anti-business shock or a pro-cannabis political climate that deters firms in general (Glaeser and Gottlieb, 2008; Moretti, 2010). The one margin that moves is the siting of the industry whose location decision is governed by the price of electricity, exactly the industry that grid lock-in predicts. With labor competition absent and a land, capital, or knowledge channel unlikely to be first order, grid lock-in through the local electricity market is the operative channel.

6 Location Choice of Data Centers

6.1 Model and Estimation

The reduced-form evidence does not depend on the structural model. The model serves only to translate the estimated electricity-cost shock into spatial reallocation and welfare magnitudes. A back-of-envelope calculation shows why a price effect of this size moves siting. A 100 megawatt facility consumes 800 gigawatt-hours a year, as Section 2.1 documents, and at an industrial rate of 8 cents per kilowatt-hour its annual electricity bill is 64 million dollars. The within-county industrial price increase of 7 percent in the most heavily cultivating

counties therefore adds 4-5 million dollars a year, every year over a facility's multi-decade life. Against the thin margins on which siting is decided, a cost increase of this size is enough to move a marginal facility to a county where electricity stays cheap.

We embed the reduced-form findings of Sections 4 and 5 in a structural discrete-choice model of data center siting. The model treats each prospective data center as a single discrete choice over a set of U.S. counties, formalizes the role of slow-moving siting fundamentals, and recovers a price coefficient that lets us compute counterfactual spatial reallocation and welfare cost under recreational cannabis legalization.

The data underlying the model is the facility-level Aterio database described in Section 3.1. The choosers are the 2,121 data center facilities announced over 2015-2024, each choosing one county from the set of contiguous U.S. counties in the year of its announcement.[14] Conditional on the choice, the facility proceeds through the construction, activation, and possibly withdrawal stages that the Aterio data track.

Decision-maker i siting in county c receives indirect utility

$$U_{ic} = \beta_p \log p_c + \delta \log pop_c + \rho\, \widehat{v}_c + \varepsilon_{ic} \tag{6}$$

where p_c is the local retail electricity price faced by industrial customers, pop_c is county population, $\widehat{v}_c$ is the control-function term defined below, and ε_{ic} is an independent type-I extreme value shock. The choice probability takes the conditional logit form

$$\Pr(c \mid i) = \frac{\exp\big(V_c\big)}{\sum_{c' \in \mathcal{C}} \exp\big(V_{c'}\big)} \tag{7}$$

where $V_c = \beta_p \log p_c + \delta \log pop_c + \rho\, \widehat{v}_c$ is the deterministic part of utility, and $\mathcal{C}$ is the choice set of counties. The price coefficient β_p is the structural object of interest, governing how strongly a higher local electricity price deters siting. We estimate the model on a sampled choice set of fifty counties per siting, the chosen county plus forty-nine sampled non-chosen counties, which delivers consistent estimates under sampling of alternatives, and bootstrap over sitings for inference.

The conditional logit imposes independence of irrelevant alternatives. Removing the price wedge in one county therefore spreads its siting probability over the remaining counties in proportion to their existing shares. We ask where mobile capacity goes when one county becomes more expensive, and this restriction answers that question. It is also why Section 6.2 reads the counterfactual as a direction and not as a facility-by-facility relocation.

[14] Where the announcement date is missing we date the decision by the activation date, and where both are missing by the construction start date.

Identification of β_p faces the standard endogeneity of price in a choice model. Counties with high unobserved siting attractiveness draw electricity-intensive activity that raises the local price, so a naive estimate is biased toward zero. We instrument the local price with recreational cannabis legalization and implement the instrument through a control function (Petrin and Train, 2010). The first stage projects the local price on the ballot-pass-by-post interaction,

$$\log p_c = \pi_0 + \pi_p \cdot Pass_{sa} \times Post_{sta} + \mathbf{w}_c'\boldsymbol{\gamma} + v_c \tag{8}$$

where $\mathbf{w}_c$ holds log population and year indicators. Legalization raises the local retail price by 12.1 percent.[15] The residual $\widehat{v}_c$ enters the utility in equation (6) and absorbs the correlation between price and unobserved attractiveness. The identifying assumption is the one maintained throughout the paper, that close-margin ballot passage is as good as randomly assigned and shifts data center siting only through the local electricity price. Fiber backbone, substation headroom, and serviced industrial land all draw data centers and all raise the local price. They enter the price residual, and the control function removes them. The identifying variation is the price movement that ballot passage produces.[16]

Table 8 reports the estimates. The price coefficient is -5.05, with a bootstrap 95 percent interval of [-6.18, -4.01], so a higher local electricity price strongly deters data centers. Population enters positively, as larger counties attract more sitings, and the control-function term is large and precisely estimated, confirming that price is endogenous and that the naive estimate understates the price response. The price coefficient is identified from the siting choices through the cannabis control function, the same variation that drives the reduced-form crowd-out, so the model translates that variation into a price elasticity and a spatial counterfactual rather than serving as an independent check. A 12.1 percent legalization-induced price increase lowers a county's relative siting probability by 44 percent, in line with the reduction in the active data center stock that Section 4 estimates directly.

6.2 Counterfactual

The first counterfactual asks how the location of data center capacity would differ absent recreational cannabis legalization. We measure capacity as operating capacity at active facilities, which totals 36 gigawatts nationally in 2024 and aligns with published U.S. data

[15] The first stage draws on the full panel and is identified largely off the early ballots, which have the most resolved post-periods and precede the satellite cultivation measure. Unlike the panel design of Section 5, the projection carries no county fixed effects, so it prices the cross-county level differences the panel absorbs, which is why it exceeds the within-county average effect of Section 5.2.

[16] Section 4.4 shows that passage does not move the other regulatory costs of doing business that could shift siting on their own.

center power demand (Shehabi et al., 2024). We build the counterfactual price in each ballot-passing county by removing the legalization effect, $\log p_c^{\text{cf}} = \log p_c - \widehat{\pi}_p \cdot Pass_{sa} \times Post_{sta}$, recompute the choice probabilities of equation (7) at the counterfactual price, and reallocate capacity to counties in proportion to the change. The model holds aggregate capacity fixed and reallocates capacity across locations, which captures the relocation of mobile capital that the locational flexibility of hyperscale facilities permits but not the reduction in total entry the reduced form also documents, so it isolates the relocation channel.

The model reallocates capacity away from the ballot-passing states toward states that never passed a recreational ballot. Appendix Figure B3 maps the direction of the state-level shift, with Texas, Pennsylvania, and Illinois the largest gainers and the ballot-passing states with the largest existing base the largest losers.[17] We read the counterfactual as the direction of the reallocation rather than a national capacity total. The discontinuity identifies the price response at the legalization margin, and a national aggregate would extrapolate it beyond the near-threshold ballots that identify the effect.

The relocation channel the model isolates redistributes fixed investment and its tax base, electricity load, and economic activity away from ballot-passing states and toward states that never passed a recreational ballot.

A reallocation of this kind should leave a trace in trade, since data center capacity is built largely from imported equipment. We look for it outside the model, in state trade in the products data centers are built from, classified by Waugh (2026). No trade variable enters the estimation. Ballot-passing states import less of the compute hardware at the core of the build-out, the imported graphics processors, servers, and storage, relative to their other trade, and the gap holds under the wild cluster bootstrap of Section 4.4. It is identified more sharply when the measure narrows to servers alone, the single most data-center-specific import, in Appendix Table C21. The broad classification, which folds in the electrical, cooling, and building materials that are sourced locally, moves nothing, and neither do exports, which record the equipment the United States sells abroad rather than what a data center installs. State trade is coarse, reported at the six-digit level and by state of destination, but its one data-center-specific margin moves the way the counterfactual predicts.

6.3 Welfare

The price wedge falls on electricity consumers in ballot-passing states who cannot relocate. Industrial and commercial users tied to local production, residential households, and public

[17] None of the three gaining states legalized through a ballot, so none carries the price wedge the counterfactual removes.

facilities pay the higher post-legalization retail rate without the tax-base and local producer-surplus gains that the data centers would have generated had they stayed. The annual consumer-surplus loss is the price wedge applied to captive 2024 consumption,

$$\Delta \mathrm{CS}_c = -\left(p_c - p_c^{\mathrm{cf}}\right) \cdot q_c^{\mathrm{captive}} \tag{9}$$

where q_c^{captive} is 2024 county electricity consumption net of the mobile data center share. The wedge falls on captive consumers in the near-threshold ballot-passing states that the design identifies, a cost borne in those states rather than by the infra-marginal adopters. It counts only the retail-price wedge and excludes the property-tax base, local employment, and agglomeration economies that the reallocated capacity would have anchored. Applied to captive 2024 consumption in the counties whose states fall inside the estimation bandwidth, the wedge costs 15.6 billion dollars a year across 648 counties, or 24.1 million dollars per county. The loss concentrates in the largest ballot-passing electricity markets, shown by state in Appendix Figure B4. The forgone investment is itself mostly imported. The 160-320 million dollars of data-center capacity a ballot-passing county loses is a capital outlay dominated by power and computing equipment, which Waugh (2026) shows is overwhelmingly imported. A county therefore loses the on-site installation, the construction, and the property-tax base the facility anchors, and not the value of the imported hardware.

7 Conclusion

Whichever electricity-intensive industry claims scarce local grid capacity first can lock it in and foreclose the next that shares the same grid. We document this with recreational cannabis legalization, which commits local capacity to indoor cultivation and forecloses data center entry. After a ballot passes, the active data center stock falls by 0.702 facilities per county and operating capacity by 16.2 MW per county; a one-square-kilometer increase in cultivation raises local retail prices 2.2 percent on average and 2.6 percent for industrial users, and electricity quantities stay flat while competition for shared labor is absent and zoning keeps the two industries off the same land. The crowd-out is anticipatory rather than contemporaneous, with developers responding to an expected resource constraint on the long horizons over which siting is decided, the defining feature of grid lock-in. The retreat reaches counties on the same grid in other states. Developers are responding to the grid they expect, not to the bill they pay.

The location-choice model puts the displacement in spatial terms that the reduced form cannot. The model reallocates capacity away from ballot-passing states toward states that

never passed a recreational ballot, and captive consumers in those ballot-passing states bear the welfare cost of the higher retail price they cannot escape. The binding spatial allocator is grid capacity, not labor, land, capital, or knowledge, and its slow adjustment converts a localized demand shock into a durable redistribution of mobile capital.

The policy lever this identifies is specific. Because the welfare cost falls on captive residential and immobile industrial consumers while the relocating data centers escape it, transmission expansion in capacity-constrained regions and retail-rate protection for captive consumers address the spatial misallocation and its distributional cost. A grid built for historical industrial demand will otherwise keep steering emerging high-load industries away from the regions whose captive consumers bear the loss.

The mechanism is not specific to cannabis or to AI, but it generalizes only where three conditions hold. Local electricity supply must adjust slowly, the entrant must consume a meaningful share of marginal grid capacity, and the foreclosed industry must be mobile and sensitive to electricity cost. Where these hold, any movable electricity-intensive entrant on a slow-adjusting grid can foreclose another, so cryptocurrency mining, semiconductor fabrication, battery manufacturing, and electrolyzer siting should compete for grid headroom in the same way, and the geography of AI compute will be shaped not only by AI demand itself but also by these rival loads. The satellite cultivation measure we build to detect one such load is validated against cannabis ground truth, and we conjecture that a similar approach could observe other localized electricity-demand shocks wherever administrative data are thin. Slow-moving electricity infrastructure can make early electricity-intensive entrants shape the geography of later mobile industries.

References

Acemoglu, Daron and Pascual Restrepo, "The Race Between Man and Machine: Implications of Technology for Growth, Factor Shares, and Employment," *American Economic Review*, 2018, *108* (6), 1488–1542.

_ and _, "Robots and Jobs: Evidence from US Labor Markets," *Journal of Political Economy*, 2020, *128* (6), 2188–2244.

Adda, Jérôme, Brendon McConnell, and Imran Rasul, "Crime and the Depenalization of Cannabis Possession: Evidence from a Policing Experiment," *Journal of Political Economy*, 2014, *122* (5), 1130–1202.

Allcott, Hunt and Daniel Keniston, "Dutch Disease or Agglomeration? The Local Economic Effects of Natural Resource Booms in Modern America," *Review of Economic Studies*, 2018, *85* (2), 695–731.

Anderson, D. Mark, Benjamin Hansen, and Daniel I. Rees, "Medical Marijuana Laws, Traffic Fatalities, and Alcohol Consumption," *Journal of Law and Economics*, 2013, *56* (2), 333–369.

Arthur, W. Brian, "Competing Technologies, Increasing Returns, and Lock-In by Historical Events," *Economic Journal*, 1989, *99* (394), 116–131.

Asher, Sam and Paul Novosad, "Rural Roads and Local Economic Development," *American Economic Review*, 2020, *110* (3), 797–823.

Autor, David H., David Dorn, and Gordon H. Hanson, "The China Syndrome: Local Labor Market Effects of Import Competition in the United States," *American Economic Review*, 2013, *103* (6), 2121–2168.

Black, Dan, Terra McKinnish, and Seth Sanders, "The Economic Impact of the Coal Boom and Bust," *Economic Journal*, 2005, *115* (503), 449–476.

Bloom, Nicholas, Mark Schankerman, and John Van Reenen, "Identifying Technology Spillovers and Product Market Rivalry," *Econometrica*, 2013, *81* (4), 1347–1393.

Borenstein, Severin and James Bushnell, "The US Electricity Industry After 20 Years of Restructuring," *Annual Review of Economics*, 2015, *7*, 437–463.

Brynjolfsson, Erik, Daniel Rock, and Chad Syverson, "The Productivity J-Curve: How Intangibles Complement General Purpose Technologies," *American Economic Journal: Macroeconomics*, 2021, *13* (1), 333–372.

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

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

_ **, Jonah B. Gelbach, and Douglas L. Miller**, "Bootstrap-Based Improvements for Inference with Clustered Errors," *Review of Economics and Statistics*, 2008, *90* (3), 414–427.

Cattaneo, Matias D., Michael Jansson, and Xinwei Ma, "Simple Local Polynomial Density Estimators," *Journal of the American Statistical Association*, 2020, *115* (531), 1449–1455.

_ **, Rocío Titiunik, and Gonzalo Vazquez-Bare**, "Analysis of Regression-Discontinuity Designs with Multiple Cutoffs or Multiple Scores," *Stata Journal*, 2020, *20* (4), 866–891.

Caughey, Devin and Jasjeet S. Sekhon, "Elections and the Regression Discontinuity Design: Lessons from Close U.S. House Races, 1942–2008," *Political Analysis*, 2011, *19* (4), 385–408.

Cengiz, Doruk, Arindrajit Dube, Attila Lindner, and Ben Zipperer, "The Effect of Minimum Wages on Low-Wage Jobs," *Quarterly Journal of Economics*, 2019, *134* (3), 1405–1454.

Cheng, Cheng, Walter J. Mayer, and Yanling Mayer, "The Effect of Legalizing Retail Marijuana on Housing Values: Evidence from Colorado," *Economic Inquiry*, 2018, *56* (3), 1585–1601.

David, Paul A., "Clio and the Economics of QWERTY," *American Economic Review*, 1985, *75* (2), 332–337.

Donaldson, Dave, "Railroads of the Raj: Estimating the Impact of Transportation Infrastructure," *American Economic Review*, 2018, *108* (4-5), 899–934.

_ **and Adam Storeygard**, "The View from Above: Applications of Satellite Data in Economics," *Journal of Economic Perspectives*, 2016, *30* (4), 171–198.

_ **and Richard Hornbeck**, "Railroads and American Economic Growth: A "Market Access" Approach," *Quarterly Journal of Economics*, 2016, *131* (2), 799–858.

Eggers, Andrew C., Anthony Fowler, Jens Hainmueller, Andrew B. Hall, and James M. Snyder, "On the Validity of the Regression Discontinuity Design for Estimating Electoral Effects: New Evidence from Over 40,000 Close Races," *American Journal of Political Science*, 2015, *59* (1), 259–274.

Electric Power Research Institute, "Powering Intelligence: Analyzing Artificial Intelligence and Data Center Energy Consumption," White Paper 3002028905, Electric Power Research Institute, Palo Alto, CA 2024.

_ , "Powering Intelligence: Updated Scenarios of U.S. Data Center Electricity Use and Power Strategies," White Paper, Electric Power Research Institute, Palo Alto, CA 2026.

Faber, Benjamin, "Trade Integration, Market Size, and Industrialization: Evidence from China's National Trunk Highway System," *Review of Economic Studies*, 2014, *81* (3), 1046–1070.

Gavrilova, Evelina, Takuma Kamada, and Floris Zoutman, "Is Legal Pot Crippling Mexican Drug Trafficking Organisations? The Effect of Medical Marijuana Laws on US Crime," *Economic Journal*, 2019, *129* (617), 375–407.

Gelman, Andrew and Guido W. Imbens, "Why High-Order Polynomials Should Not Be Used in Regression Discontinuity Designs," *Journal of Business & Economic Statistics*, 2019, *37* (3), 447–456.

Glaeser, Edward L. and Joshua D. Gottlieb, "The Economics of Place-Making Policies," *Brookings Papers on Economic Activity*, 2008, *39* (1), 155–239.

Goldfarb, Avi and Catherine Tucker, "Digital Economics," *Journal of Economic Literature*, 2019, *57* (1), 3–43.

Greenstone, Michael, Richard Hornbeck, and Enrico Moretti, "Identifying Agglomeration Spillovers: Evidence from Winners and Losers of Large Plant Openings," *Journal of Political Economy*, 2010, *118* (3), 536–598.

Grembi, Veronica, Tommaso Nannicini, and Ugo Troiano, "Do Fiscal Rules Matter?," *American Economic Journal: Applied Economics*, 2016, *8* (3), 1–30.

Henderson, J. Vernon, Adam Storeygard, and David N. Weil, "Measuring Economic Growth from Outer Space," *American Economic Review*, 2012, *102* (2), 994–1028.

Hsieh, Chang-Tai and Enrico Moretti, "Housing Constraints and Spatial Misallocation," *American Economic Journal: Macroeconomics*, 2019, *11* (2), 1–39.

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

Kahn, Matthew E. and Erin T. Mansur, "Do Local Energy Prices and Regulation Affect the Geographic Concentration of Employment?," *Journal of Public Economics*, 2013, *101*, 105–114.

Lee, David S., "Randomized Experiments from Non-Random Selection in U.S. House Elections," *Journal of Econometrics*, 2008, *142* (2), 675–697.

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

McCrary, Justin, "Manipulation of the Running Variable in the Regression Discontinuity Design: A Density Test," *Journal of Econometrics*, 2008, *142* (2), 698–714.

Mills, Evan, "The Carbon Footprint of Indoor Cannabis Production," *Energy Policy*, 2012, *46*, 58–67.

Moretti, Enrico, "Local Multipliers," *American Economic Review*, 2010, *100* (2), 373–377.

_ , "The Effect of High-Tech Clusters on the Productivity of Top Inventors," *American Economic Review*, 2021, *111* (10), 3328–3375.

Mytton, David, "Data Centre Water Consumption," *npj Clean Water*, 2021, *4* (1), 11.

Petrin, Amil and Kenneth Train, "A Control Function Approach to Endogeneity in Consumer Choice Models," *Journal of Marketing Research*, 2010, *47* (1), 3–13.

PJM Interconnection, "2025/2026 Base Residual Auction Report," Technical Report, PJM Interconnection, Audubon, Pennsylvania 2024.

Rambachan, Ashesh and Jonathan Roth, "A More Credible Approach to Parallel Trends," *Review of Economic Studies*, 2023, *90* (5), 2555–2591.

Rand, Joseph, Nick Manderlink, Will Gorman, Ryan Wiser, Joachim Seel, Julie Mulvaney Kemp, Seongeun Jeong, and Fritz Kahrl, "Queued Up: 2024 Edition, Characteristics of Power Plants Seeking Transmission Interconnection As of the End of 2023," Technical Report LBNL-2001637, Lawrence Berkeley National Laboratory, Berkeley, California 2024.

Serrato, Juan Carlos Suárez and Owen Zidar, "Who Benefits from State Corporate Tax Cuts? A Local Labor Markets Approach with Heterogeneous Firms," *American Economic Review*, 2016, *106* (9), 2582–2624.

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.

Siddik, Md Abu Bakar, Arman Shehabi, and Landon Marston, "The Environmental Footprint of Data Centers in the United States," *Environmental Research Letters*, 2021, *16* (6), 064017.

Summers, Hailey M., Evan Sproul, and Jason C. Quinn, "The Greenhouse Gas Emissions of Indoor Cannabis Production in the United States," *Nature Sustainability*, 2021, *4*, 644–650.

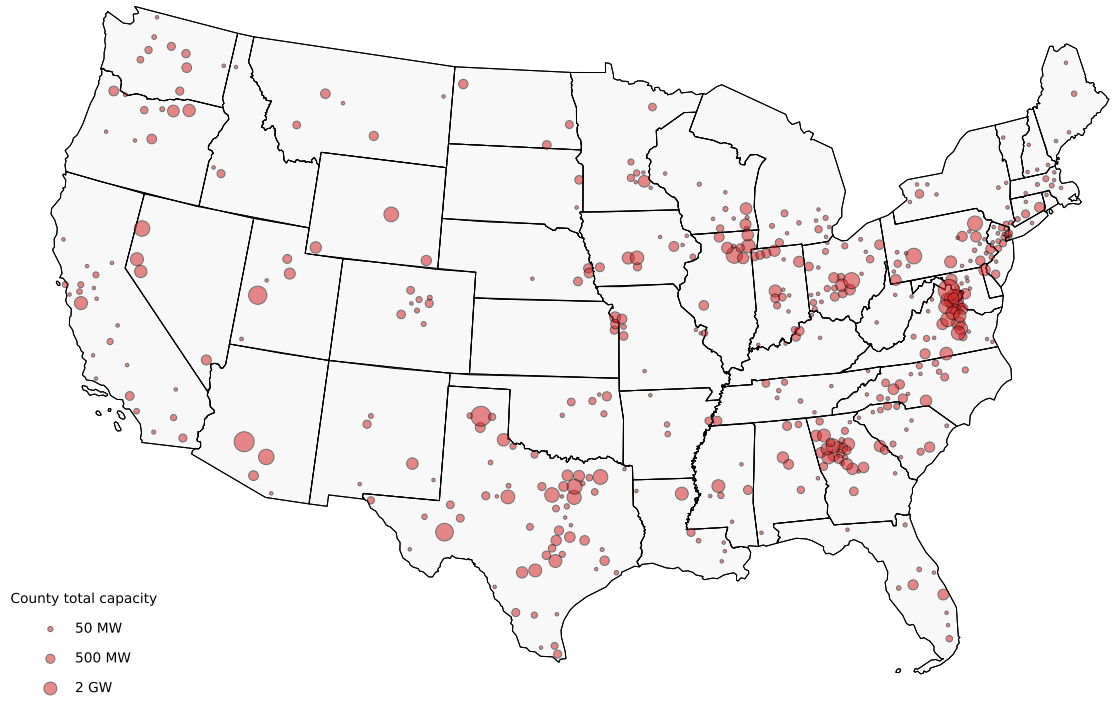
Turner and Townsend, "Data Centre Construction Cost Index 2025-2026," Technical Report, Turner and Townsend, London 2025.

Unruh, Gregory C., "Understanding Carbon Lock-in," *Energy Policy*, 2000, *28* (12), 817–830.

Waugh, Michael E., "Trade in AI-Related Products," Working Paper 35053, National Bureau of Economic Research 2026.

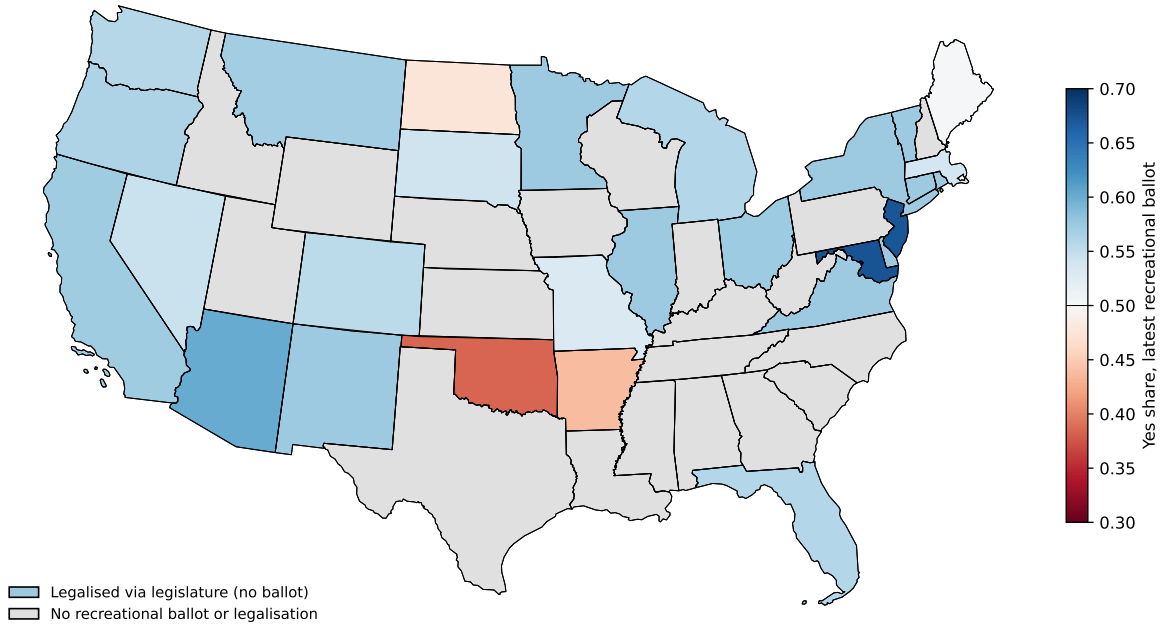
Wen, Hefei, Jason M. Hockenberry, and Janet R. Cummings, "The Effect of Medical Marijuana Laws on Adolescent and Adult Use of Marijuana, Alcohol, and Other Substances," *Journal of Health Economics*, 2015, *42*, 64–80.

Figure 1. U.S. data center locations



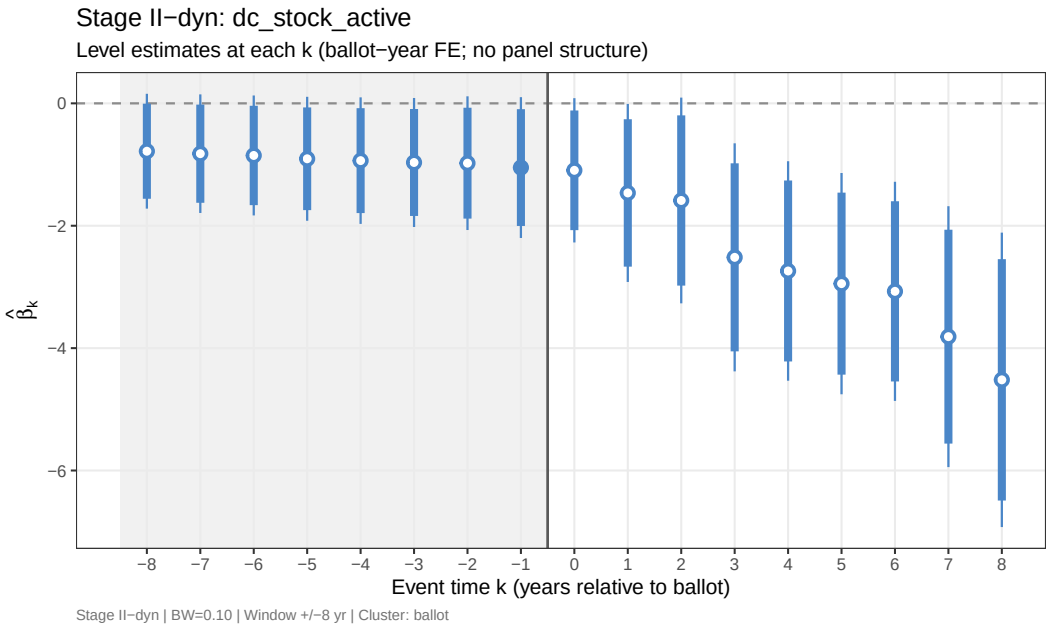
Notes: Each red marker is a county centroid sized by the total nameplate power capacity of data center facilities located in that county. The sample includes 5,140 facilities across 461 counties and pools all development stages, from announcement to operation. Marker area is proportional to the square root of total county capacity in megawatts. Continental U.S. shown, with Alaska and Hawaii omitted from the map. Data source: Aterio facility-level data center database.

Figure 2. State outcomes on the latest recreational cannabis ballot



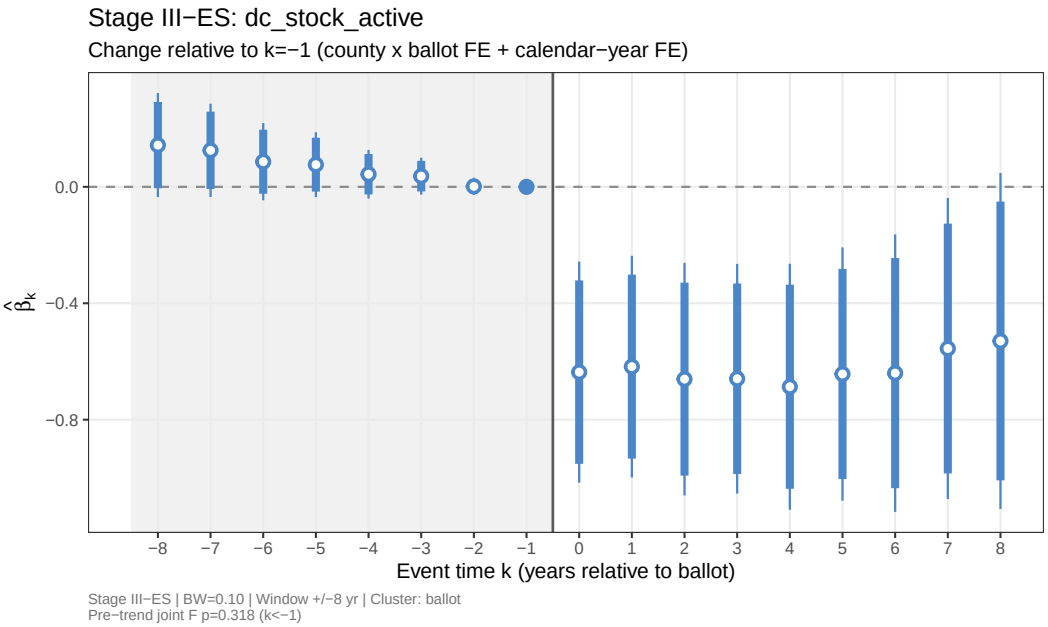
Notes: States are shaded by the Yes share on the most recent recreational cannabis ballot, with passing ballots taking precedence over later defeats. The diverging color scale is centered at the 50 percent passage threshold, so blue shades indicate states that legalized by ballot and red shades indicate states that rejected. States that legalized recreational use through legislative action rather than ballot are shaded light blue and labeled separately. States with no recreational ballot and no recreational legalization are shaded gray. Continental U.S. shown, with Alaska and Hawaii omitted from the map. Data source: Ballotpedia recreational cannabis ballot history.

Figure 3. Cross-sectional data center stock gap by event time



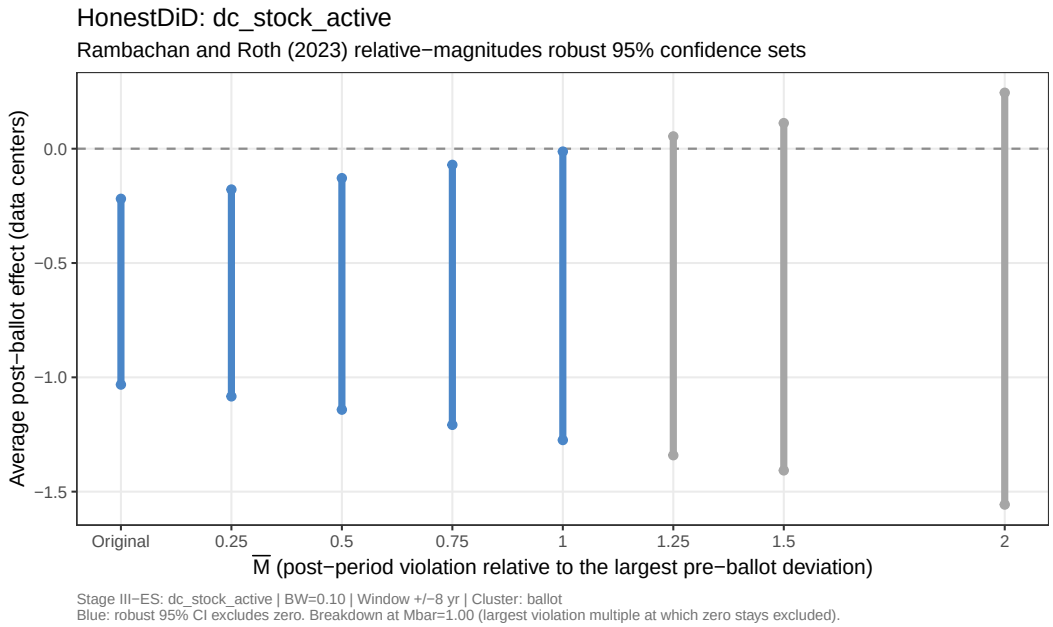
Notes: The figure plots the pooled cross-sectional RD coefficient on *Pass*, estimated separately at each event time k. The pass-fail gap in the active data center stock is present before the ballot and widens after passage. The 95% confidence intervals are based on standard errors clustered at the ballot level. The estimation sample is restricted to $|R_{sa}| \leq 0.10$. The outcome has a mean of 0.446 active facilities per county-year.

Figure 4. Data center stock around cannabis legalization



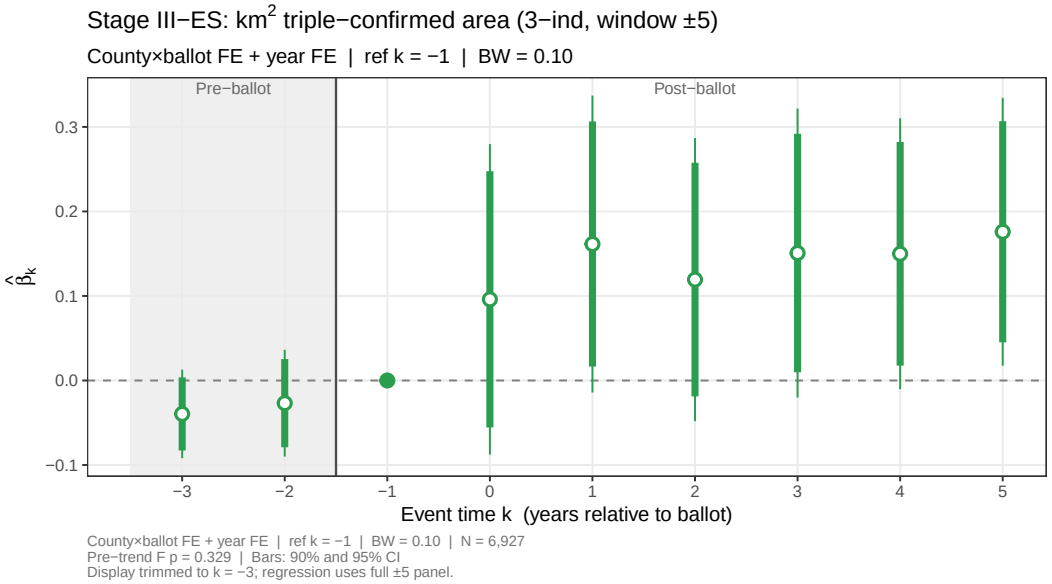
Notes: The figure reports a panel difference-in-discontinuities event study in which the single post-ballot term is replaced by interactions of *Pass* with event time, and the year before the ballot is the omitted reference period. All specifications include county-by-ballot and calendar-year fixed effects. The pre-ballot coefficients are jointly indistinguishable from zero. The pass-fail gap opens at $k = 0$ and persists through $k = +8$. The 95% confidence intervals are based on standard errors clustered at the ballot level. The estimation sample is restricted to $|R_{sa}| \leq 0.10$. The outcome has a mean of 0.446 active facilities per county-year.

Figure 5. Sensitivity of the post-ballot effect to violations of parallel trends



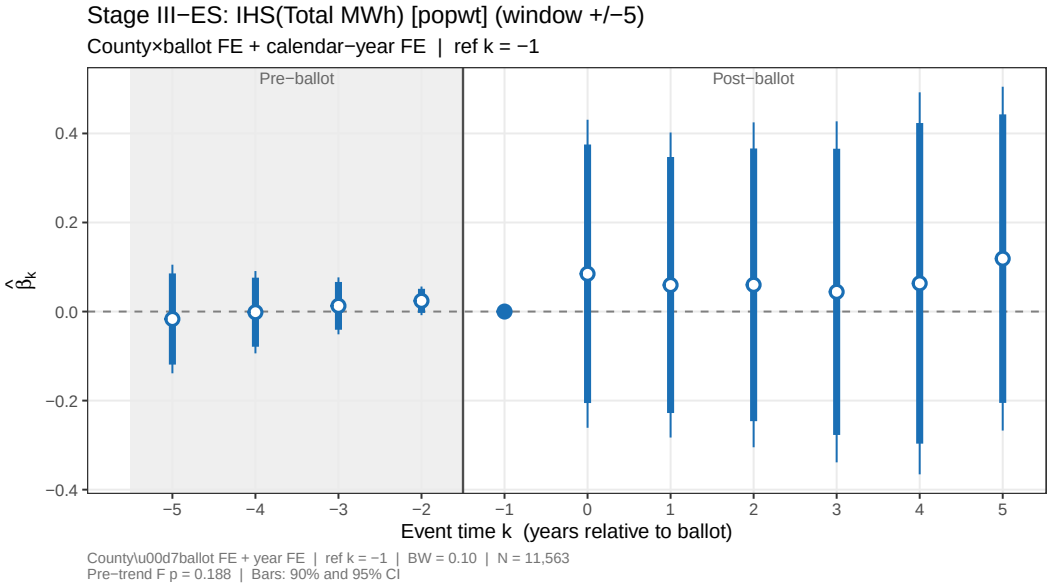
Notes: The figure reports the Rambachan and Roth (2023) relative-magnitudes robust 95% confidence set for the average post-ballot effect on the active data center stock, as the permitted post-period violation of parallel trends grows from zero (the original confidence interval, at $\bar{M} = 0$) to twice the largest violation observed across the seven pre-ballot years ($\bar{M} = 2$). The average runs over the nine post-ballot event-study coefficients and equals a 0.63-facility decline, close to the 0.70 static estimate. Intervals that exclude zero are shown in blue and those that admit zero in gray. The robust set excludes zero through $\bar{M} = 1.00$, a post-period violation as large as the largest deviation across the pre-ballot years, and first admits zero at $\bar{M} = 1.25$. The event study is the Stage III specification of Figure 4, with standard errors clustered at the ballot level and the sample restricted to $|R_{sa}| \leq 0.10$.

Figure 6. Detected cannabis cultivation around legalization



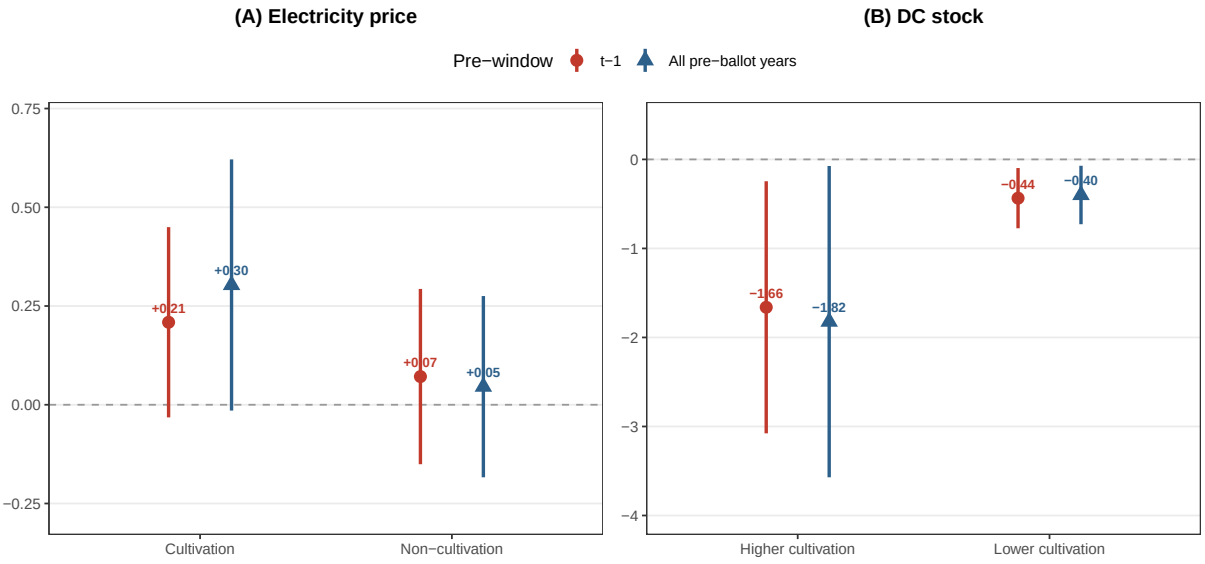
Notes: The figure reports a panel difference-in-discontinuities event study for three-indicator triconfirmed cultivation area in square kilometers, estimated over a plus or minus five-year window, and the year before the ballot is the omitted reference period. The 95% confidence intervals are based on standard errors clustered at the ballot level. The estimation sample is restricted to $|R_{sa}| \leq 0.10$. The outcome has a mean of 0.047 square kilometers per county-year.

Figure 7. Total electricity consumption around legalization



Notes: The figure reports a panel difference-in-discontinuities event study for population-weighted total electricity consumption in log megawatt-hours, estimated over a plus or minus five-year window, and the year before the ballot is the omitted reference period. The pre-trends are flat, and the post-ballot point estimates drift modestly positive but are not distinguishable from zero. The 95% confidence intervals are based on standard errors clustered at the ballot level. The estimation sample is restricted to $|R_{sa}| \leq 0.10$. The plus or minus eight-year version appears in the appendix as Figure B2. The outcome has a mean of 1.43 million megawatt-hours per county-year.

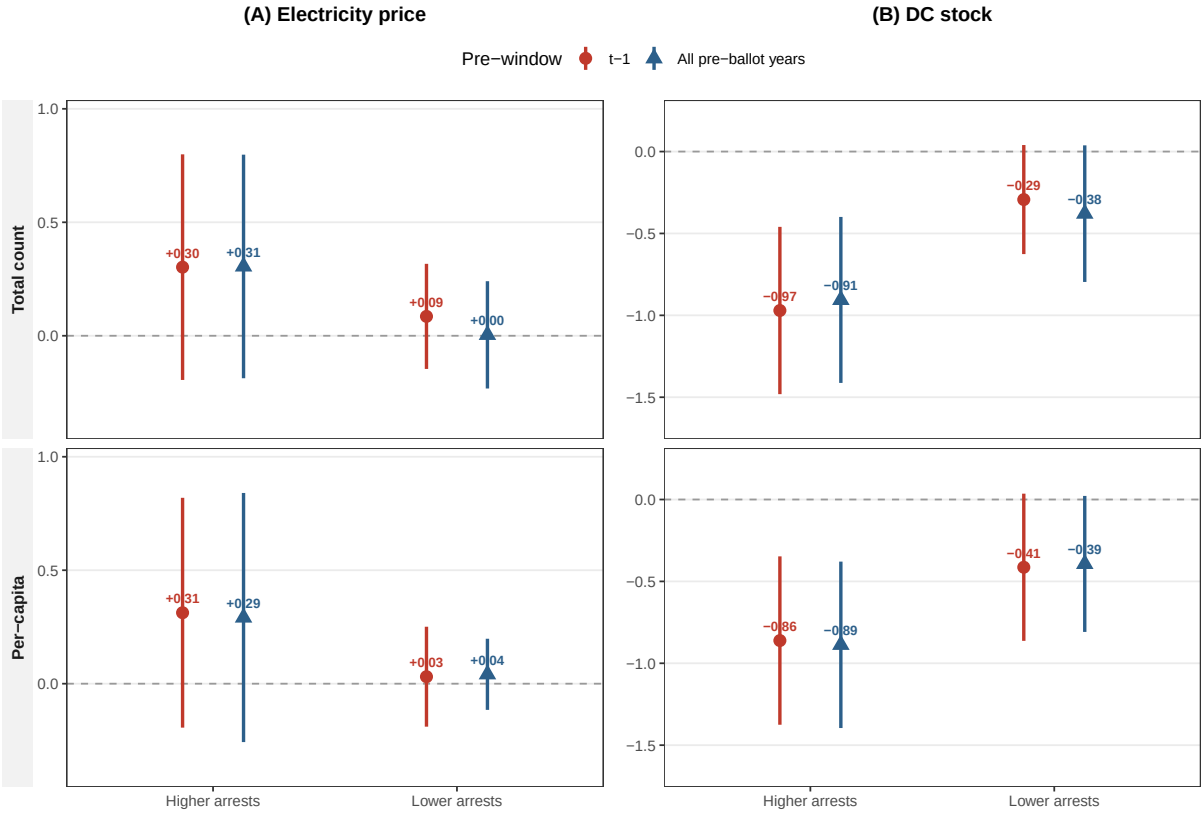
Figure 8. Price and data center responses by pre-ballot cultivation exposure



County sample split by pre−ballot cannabis cultivation status. Each column has its own y−axis: (A) electricity price uses the three−indicator any−detection classifier with the +/−3 estimation window; (B) data−center stock uses the two−indicator mean−cut classifier with the +/−8 estimation window. Within each cell, pre−window = t−1 (red circles) vs mean over all pre−ballot years (blue triangles). Bandwidth |R_sa| <= 0.10; 95% CI from the RD−DiD's own interval (cluster = ballot, fixest confint), not +/− 1.96.

Notes: The figure reports the difference-in-discontinuities Pass × Post effect on the average retail electricity price (Panel A, log) and the active data center stock (Panel B, levels), with counties split by pre-ballot satellite-detected cultivation status. Panel A uses the three-indicator any-detection cultivation classifier with the plus or minus three-year estimation window. Panel B uses the two-indicator mean-cut classifier with the plus or minus eight-year estimation window. The two series within each panel use the cultivation classifier in the year before the ballot ($t-1$) and the lenient mean over all pre-ballot years. Both the price increase and the data center retreat concentrate in counties with pre-ballot cultivation. The bandwidth is $|R_{sa}| \leq 0.10$, and the year before the ballot is the omitted reference period. All specifications include county-by-ballot and calendar-year fixed effects. The confidence intervals are the estimator's own intervals with standard errors clustered at the ballot level.

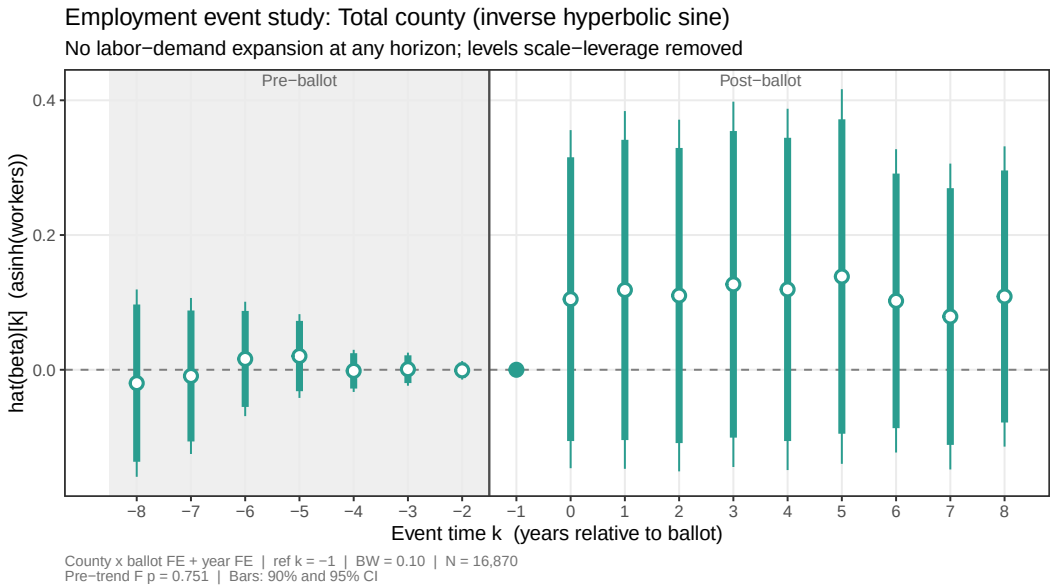
Figure 9. Price and data center responses by pre-ballot cannabis-supply enforcement



Rows: total−count splitter (primary; mechanism = absolute supply) vs per−capita splitter (robustness). Each column has its own y−axis, shared between the two splitter rows. Within each cell: pre−window = t−1 (red circles) vs mean over all pre−ballot ASR years (blue triangles, lenient). Global median split within each splitter; |R_sa| <= 0.10, |tau| <= 8. 95% CI from the RD−DiD's own interval (cluster = ballot, fixest confint), not +/− 1.96.

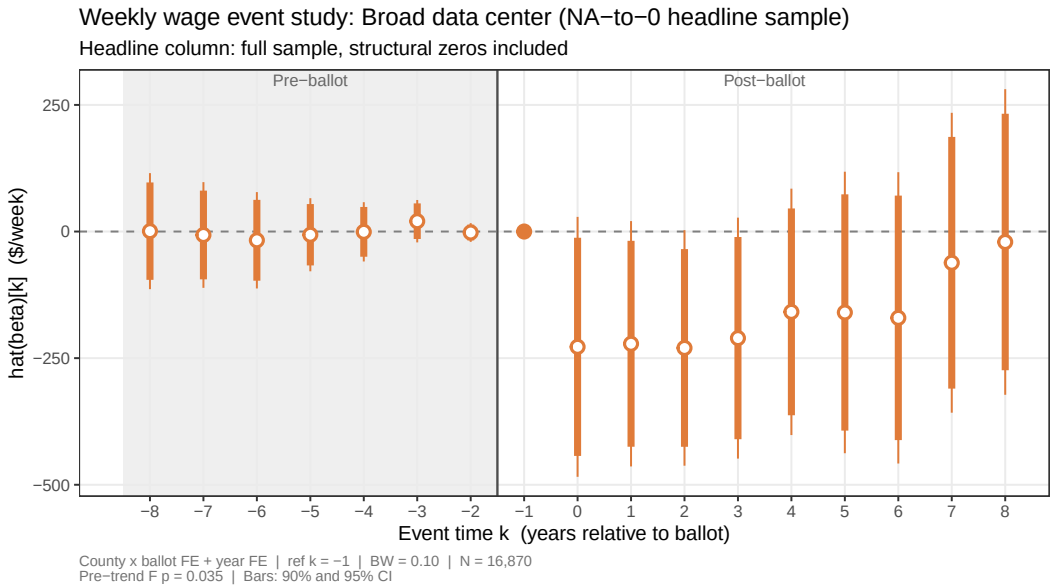
Notes: The figure reports the difference-in-discontinuities Pass × Post effect on the average retail electricity price (Column A, log) and the active data center stock (Column B, levels), with counties split by pre-ballot cannabis-supply enforcement, namely marijuana sale and manufacturing arrests from the UCR Arrests by Age, Sex, and Race series (Kaplan, openICPSR #102263). Unlike the satellite cultivation measure in Figure 8, the arrest measure has coverage back to the 1970s and so classifies all 26 ballots, including the early heavy-cultivator ballots of Colorado and Washington in 2012 that satellite detection cannot see before 2017. The top row uses the total count of pre-ballot sale and manufacturing arrests as the splitter, the mechanism-relevant margin, since what matters for resource competition is the absolute amount of cannabis supply and electricity consumption, not the per-capita rate. The bottom row uses the per-capita rate as a robustness check that high total arrests do not merely reflect a populous county. Counties are split at the sample-wide median within each splitter, and the two series within each cell use the splitter in the year before the ballot ($t-1$) and the lenient mean over all pre-ballot years. The data center retreat is larger in high-arrest counties under the total-count splitter, a difference of -0.68 active centers at the $t-1$ split, and the qualitative pattern holds under the per-capita splitter at -0.45. The price increase is also larger in high-arrest counties under both splitters, though the difference across groups is not distinguishable. The estimation window is plus or minus eight years, the bandwidth is $|R_{sa}| \leq 0.10$, and the year before the ballot is the omitted reference period. All specifications include county-by-ballot and calendar-year fixed effects. The confidence intervals are the estimator's own intervals with standard errors clustered at the ballot level.

Figure 10. Total county employment around legalization



Notes: The figure reports a panel difference-in-discontinuities event study for total county employment in inverse hyperbolic sine, estimated over a plus or minus eight-year window, and the year before the ballot is the omitted reference period. The 95% confidence intervals are based on standard errors clustered at the ballot level. The estimation sample is restricted to $|R_{sa}| \leq 0.10$. The estimated path is flat before and after the ballot, a precise null around legalization. Total county employment has a pre-period mean of 57,136 and a median of 6,701 workers per county-year in the estimation sample (the gap reflects the right skew from large urban counties such as Los Angeles). The outcome is constructed with NA filled to zero to match the inverse-hyperbolic-sine transform. See Table 7.

Figure 11. Data center sector wages around legalization



Notes: The figure reports a panel difference-in-discontinuities event study for data-center-sector wages with non-operating county-years set to zero, estimated over a plus or minus eight-year window, and the year before the ballot is the omitted reference period. The 95% confidence intervals are based on standard errors clustered at the ballot level. The estimation sample is restricted to $|R_{sa}| \leq 0.10$. The outcome has a mean of 1,817 dollars per week among operating facilities.

Table 1. Electricity prices and detected cultivation within county

	Average (1)	Commercial (2)	Industrial (3)
$\hat{\gamma}$: inverse hyperbolic sine(triconf km^2)	0.0508***	0.0457***	0.0700***
	(0.0071)	(0.0134)	(0.0074)
Fixed effects			
County × ballot	Yes	Yes	Yes
Calendar year	Yes	Yes	Yes
Fit statistics			
Observations	6,868	6,868	6,841
R^2	0.84735	0.85561	0.89380
Within R^2	0.00585	0.00443	0.00514

Notes: The table reports within-county estimates of the association between detected cultivation and electricity prices, at the county-year level. The outcome in every column is the natural log of the population-weighted retail electricity price in dollars per kWh for a customer class, namely the all-customer average in Column (1), commercial in Column (2), and industrial in Column (3). The regressor is the inverse hyperbolic sine of triple-confirmed satellite cultivation area. All specifications include county-by-ballot and calendar-year fixed effects. The estimation bandwidth is $|R_{sa}| \leq 0.10$, and the sample is restricted to 2017-2024. Standard errors are clustered at the ballot level, with 26 or fewer clusters, and are reported in parentheses. Three, two, and one asterisks denote statistical significance at the 1, 5, and 10 percent levels, respectively.

Table 2. Cannabis cultivation, electricity prices, and consumption

	Total		Commercial		Industrial	
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A. Electricity price						
$\hat{\beta}_1$: Pass×Post	0.1199	0.1197	0.1823	0.1823	0.1449	0.1444
	(0.1064)	(0.1051)	(0.1341)	(0.1331)	(0.1195)	(0.1179)
$\hat{\beta}_2$: km^2×Pass×Post		0.0219***		0.0231***		0.0264*
		(0.0072)		(0.0066)		(0.0136)
Observations	6,868	6,868	6,868	6,868	6,841	6,841
Panel B. Electricity consumption						
$\hat{\beta}_1$: Pass×Post	0.0485	0.0492	-0.0579	-0.0574	0.0108	0.0114
	(0.3660)	(0.3650)	(0.3940)	(0.3930)	(0.2085)	(0.2078)
$\hat{\beta}_2$: km^2×Pass×Post		0.0295		0.0369**		0.0266
		(0.0170)		(0.0163)		(0.0220)
Observations	6,868	6,868	6,868	6,868	6,868	6,868
km^2 interaction	No	Yes	No	Yes	No	Yes

Notes: The table reports panel difference-in-discontinuities estimates of the effect of recreational cannabis legalization on electricity prices and consumption, at the county-year level. All specifications include county-by-ballot and calendar-year fixed effects. The estimation bandwidth is $|R_{sa}| \leq 0.10$, and the event window is plus or minus five years around the ballot. The sample is restricted to 2017-2024, the period over which the three-indicator satellite measure is available. Standard errors are clustered at the ballot level, with 26 or fewer clusters, and are reported in parentheses. The "Total" columns report the average retail price in Panel A and total consumption in Panel B. Price outcomes are in logs and consumption outcomes are in the inverse hyperbolic sine of MWh. Odd columns omit the interaction and even columns add $\hat{\beta}_2$, the marginal effect per km^2 of detected cannabis in a legalized county post-ballot. Three, two, and one asterisks denote statistical significance at the 1, 5, and 10 percent levels, respectively.

Table 3. Sample composition of the utility-grid and balancing-authority (BA)-grid spillover footprints

Diagnostic	Utility grid	BA grid
Footprint coverage		
Footprint counties	1,173	2,429
Distinct grid units	138	35
Ballots	26	26
Panel rows (county-ballot-year)	38,556	192,933
Spatial composition		
Share in ballot state	54%	11%
Share out of ballot state	46%	89%
Cultivation dose		
Mean own cultivation (km^2)	0.024	0.019
Mean rest-of-grid cultivation (km^2)	0.859	11.413
Treated group counts (post-ballot)		
G1 in-state cultivating (county-ballots)	112	112
G2 in-state non-cultivating (county-ballots)	609	609
G3 cross-state grid-mate (county-ballots)	386	4,177

Notes: Each grid footprint is constructed for the 26 ballots in $|R_{sa}| \leq 0.10$ as the union of counties served by a grid unit (utility or balancing authority) that also serves the balloting state. The BA footprint is much wider than the utility footprint because BAs span multi-state RTO/ISO sets (such as SPP, MISO, PJM), each covering more than a thousand counties. The G1, G2, G3 partition holds the in-state cells fixed across grids by construction; the difference in G3 reflects how many cross-state grid-mate counties each grid definition pulls in. The mean rest-of-grid cultivation also scales with the size of the grid-unit footprint, because cultivation is summed over more counties at the BA level.

Table 4. Spillover of cannabis cultivation onto county electricity price (continuous dose, RD-DiD, ±8 yr)

Grid	Utility		BA	
Spillover	pooled	split	pooled	split
Model:	(1)	(2)	(3)	(4)
Variables				
Pass×Post	0.0485	0.0678	-0.0137	-0.0141
	(0.0912)	(0.0853)	(0.0231)	(0.0241)
own km^2×P×P	0.0072*	0.0037	0.0040	0.0040
	(0.0039)	(0.0040)	(0.0029)	(0.0030)
spillover km^2 (pooled)	0.0046*		0.0005**	
	(0.0024)		(0.0002)	
within-state km^2		0.0090***		0.0004
		(0.0031)		(0.0003)
cross-state km^2		0.0014		0.0005**
		(0.0023)		(0.0002)
Fixed-effects				
county_fips-ballot_id	Yes	Yes	Yes	Yes
year	Yes	Yes	Yes	Yes
Fit statistics				
Observations	15,926	15,926	88,561	88,561

Clustered (ballot_id) standard-errors in parentheses
*Signif. Codes: ***: 0.01, **: 0.05, *: 0.1*

Outcome: log population-weighted average retail price. Continuous dose: own cultivation (km^2) and rest-of-grid cultivation, each interacted with Pass×Post; the "split" columns separate rest-of-grid into within-state and cross-state components. Columns 1-2 use the utility-grid footprint (all counties on a utility that also serves the balloting state); columns 3-4 use the BA-grid footprint (all counties on a balancing authority that also serves the balloting state). Sample composition is summarized in Table 3. Sample window: 2017-2024 (3-ind cultivation coverage). County×ballot and year FE; BW $|R_{sa}| \leq 0.10$; SE clustered by ballot. $^{*}p < .1, ^{**}p < .05, ^{***}p < .01$.

Table 5. Spillover of cannabis cultivation onto county data-center stock (RD-DiD, ±8 yr)

Grid	Utility		BA	
Spillover	pooled	split	pooled	split
Model:	(1)	(2)	(3)	(4)
Variables				
G1 direct (legal+cult)	-0.2934*	-0.2988*	-0.1361	-0.1318
	(0.1584)	(0.1661)	(0.0973)	(0.0929)
Spillover (pooled)	-0.3714**		-0.1999**	
	(0.1403)		(0.0772)	
G2 within-state		-0.3468**		-0.1724
		(0.1387)		(0.1012)
G3 cross-state		-0.4194**		-0.2009**
		(0.1741)		(0.0766)
Fixed-effects				
county_fips-ballot_id	Yes	Yes	Yes	Yes
year	Yes	Yes	Yes	Yes
Fit statistics				
Observations	29,518	29,518	134,358	134,358

Clustered (ballot_id) standard-errors in parentheses
*Signif. Codes: ***: 0.01, **: 0.05, *: 0.1*

Outcome: active data-center count (levels). Dummy groups defined as in the price spillover table: G1 in-state cultivating; G2 in-state non-cultivating; G3 other-state same-grid (non-legal); "Spillover (pooled)" = G2+G3. Columns 1-2 use the utility-grid footprint; columns 3-4 use the BA-grid footprint; sample composition is summarized in Table 3. The grid-wide crowd-out across G1, G2, and G3 is the basis for separating channels: G3's data-center retreat occurs without any cross-state retail-price response (see Table 4), which cannot be a contemporaneous price response and identifies the forward-looking anticipation channel. The within-state crowd-out is consistent with both a contemporaneous price channel and the same anticipation channel; this design does not separate the two within state. County×ballot and year FE; BW $|R_{sa}| \leq 0.10$; SE clustered by ballot.
$^{*}p < .1, ^{**}p < .05, ^{***}p < .01$.

Table 6. Within-state and cross-state spillovers on retail electricity price: estimate, 95% confidence interval, p-value, and minimum detectable effect (per km^2 of cultivation, in percent)

Grid	Spillover	Estimate (%)	SE (%)	95% CI (%)	p-value	MDE (%)
Utility	Within-state	+0.904***	0.314	[+0.253, +1.554]	0.009	0.879
Utility	Cross-state	+0.140	0.226	[-0.328, +0.608]	0.541	0.632
BA	Within-state	+0.041	0.028	[-0.016, +0.099]	0.148	0.077
BA	Cross-state	+0.049**	0.019	[+0.010, +0.087]	0.016	0.052

Notes: Continuous dose specification from Table 4 (utility and BA grids): within-state and cross-state cultivation in square kilometers, each interacted with Pass×Post, with own cultivation and the running-variable polynomial included. County×ballot and year FE; BW $|R_{sa}| \leq 0.10$; SE clustered by ballot. Coefficients are reported in percent per km^2 of cultivation (the outcome is in log electricity price, so the raw coefficient times 100 is the approximate percent change). The minimum detectable effect (MDE) is $(z_{0.975} + z_{0.80}) \times \text{SE} \approx 2.80 \times \text{SE}$, the smallest absolute effect the design would reject at $\alpha = 0.05$ two-sided with 80% power. The utility-grid within-state spillover of +0.90% per km^2 is the headline; the cross-state MDE of 0.63% per km^2 on the same grid is 70% of that headline, so the cross-state null bounds the wholesale leakage as economically small rather than the limit of statistical resolution. Under the BA grid the within-state vs cross-state contrast inverts because the BA footprint averages over many distant grid-mate counties (see Table 3). $^{*}p < .1$,$^{**}p < .05$,$^{***}p < .01$.

Table 7. Labor-market wages and employment

	Weekly wage, levels ($/wk)				Employment, inverse hyperbolic sine			
	(1) County total	(2) Data center	(3) Cannabis cultivation	(4) HVAC /elec.	(5) County total	(6) Data center	(7) Cannabis cultivation	(8) HVAC /elec.
Pass × Post	-6.2	-195.6	-70.7	-76.9	0.112	-0.054	0.301	-0.027
	(31.2)	(137.9)	(76.1)	(65.6)	(0.135)	(0.156)	(0.412)	(0.116)
Pre-period mean	745.0	240.7	52.4	683.5	57,136	100.9	19.7	778.3
County × ballot FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	16,870	16,870	16,870	16,870	16,870	16,870	16,870	16,870
R^2	0.932	0.745	0.483	0.849	0.966	0.902	0.689	0.958
Within R^2	0.006	0.006	0.004	0.005	0.002	0.002	0.003	0.002

Notes: Each column is a separate panel difference-in-discontinuities regression of the named outcome on the Pass × Post interaction, using the running-variable controls and fixed effects listed below. Wage in columns 1-4 is in levels in dollars per week, and employment in columns 5-8 is in the inverse hyperbolic sine, asinh(workers), which handles the true zeros without an additive constant and removes the large-county scale leverage that dominates a levels headcount regression. Missing QCEW cells are filled with zero before transformation, so a county-year with no firm or a disclosure-suppressed cell contributes $0/week or 0 workers, and the full canonical 16,870-row panel difference-in-discontinuities sample enters every column. Columns (1)/(5) are the NAICS 10 county total, where wage is the employment-weighted county-average and employment is total headcount. Columns (2)/(6) are broad data centers in NAICS 518210 + 541513, columns (3)/(7) cannabis cultivation in NAICS 111419 + 111998, and columns (4)/(8) HVAC and electrical contractors in NAICS 2382, the contractors shared by indoor-grow and data-center construction. For these sparser sectors the zero fill means the column tests firm entry as much as wage competition or labor demand. Pre-period means are in raw units: $/week for wage, workers for employment. The labor-competition channel predicts at least one wage column *rises* and that workers are pulled *into* those sectors; no column on either margin shows that positive response. Both HVAC and electrical point estimates are instead negative, consistent with a demand-side channel in which foreclosed data-center construction lowers demand for those contractors. The levels version of the employment columns, which is dominated by a few large treated counties, is reported in §5.5. All specifications include county-by-ballot and calendar-year fixed effects, and the estimation bandwidth is $|R_{sa}| \leq 0.10$. Standard errors are clustered at the ballot level, with 26 or fewer clusters, and are reported in parentheses. Three, two, and one asterisks denote statistical significance at the 1, 5, and 10 percent levels, respectively.

Table 8. Data center location-choice model

	Coefficient	95% CI
log retail electricity price	-5.05	[-6.18, -4.01]
log county population	0.99	[0.96, 1.03]
Control function, price residual	3.45	[2.37, 4.58]
Data center sitings (choosers)	2,121	
Alternatives per chooser	50	
First-stage price effect of legalization	12.1%	

Notes: The table reports conditional logit estimates of data center location choice. The choosers are the 2,121 facilities announced over 2015 to 2024, each choosing from a sampled choice set of fifty counties, the chosen county and forty-nine sampled non-chosen counties. Utility is linear in the log retail electricity price, log county population, and the first-stage price residual that enters as a control function. The local price is instrumented by recreational cannabis legalization, with a first-stage price effect of 12.1 percent. The 95 percent intervals are bootstrapped over sitings.

Online Appendix

A Cannabis Cultivation Detection

This section documents the satellite-based detection algorithm used to construct the county-year measure of indoor cannabis cultivation activity described in Section 3.3. The pipeline combines three remote-sensing signals into a composite county-year activity index. Each signal is processed independently, validated against ground-truth licensed-facility locations where available, and then merged into a confidence-weighted composite.

The first signal is monthly cloud-free composite radiance from the Visible Infrared Imaging Radiometer Suite Day/Night Band on the Suomi NPP and NOAA-20 satellites, distributed through the NASA Black Marble (VNP46) product. The Day/Night Band has a native ground sample distance of 750 meters at nadir and an effective dynamic range that resolves anthropogenic light sources two orders of magnitude below the saturation level of legacy DMSP/OLS. We extract the monthly cloud-free, lunar-corrected, atmosphere-corrected at-sensor radiance from VNP46A3, mask water and persistently bright industrial pixels using a baseline mean computed over the 2012 calendar year, and aggregate to the county-year as the sum of pixel-level radiance anomalies above the baseline. Anomaly pixels with values above a county-specific 95th percentile threshold are flagged as candidate cultivation pixels.

The second signal is at-surface brightness temperature from Landsat 8 thermal infrared bands 10 and 11, taken from the USGS Collection 2 Level-2 product at 30-meter resampled resolution. Landsat 8 has a 16-day revisit cycle at the equator and a ten-thirty local crossing time, which captures the heat plume from large-scale lighting and cooling operations during the late-morning warm-up period. We compute a deseasonalized brightness-temperature anomaly relative to a 2013-2014 baseline at the pixel level, mask cloud and cloud-shadow pixels using the QA_PIXEL band, and aggregate to the county-year as the sum of pixel-level anomalies above a fixed three-Kelvin threshold. Pixels exceeding the threshold for three or more cloud-free scenes in a year are flagged as candidate cultivation pixels.

The third signal is optical rooftop detection from Sentinel-2 surface reflectance at 10-meter resolution, distributed through the Copernicus program with a five-day revisit and available from 2017. Greenhouse and warehouse roofs are bright and spectrally flat, so we combine a brightness index across the visible and near-infrared bands with a vegetation index that separates engineered roofing from vegetated and bare ground. Pixels whose brightness exceeds a county-specific threshold while the vegetation index stays low are flagged as cultivation-consistent roofing, and the flagged rooftop area in each county-year is the third

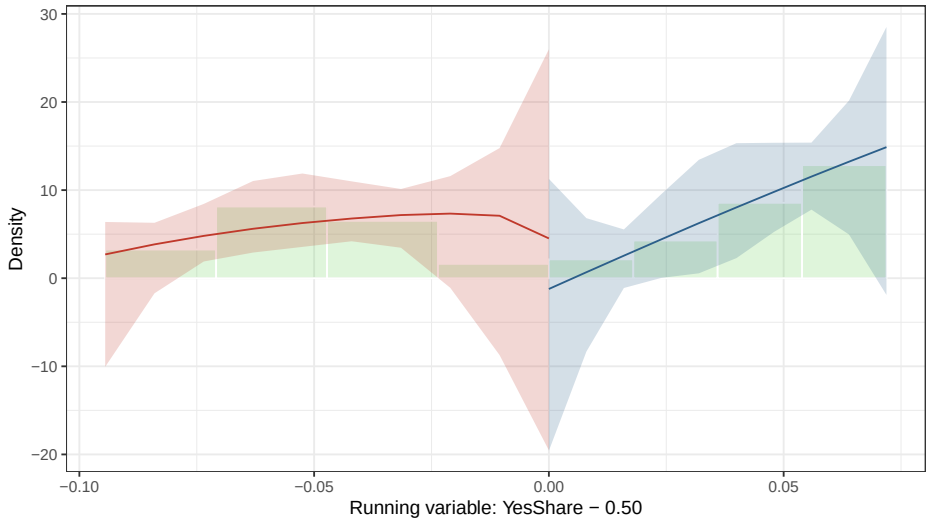
candidate signal. The Sentinel-2 record is what sets 2017 as the first year of the three-signal measure.

The three candidate signals are merged into a composite county-year activity index using inverse-variance weights estimated against a held-out validation sample of state-licensed cultivation facilities. State seed-to-sale licensing records from Colorado, Oregon, and Washington provide ground-truth cultivation locations through 2024, comprising 9,418 licensed cultivation sites at the time of the most recent extract. We geocode each licensed site to its county and year of license issuance and use this geo-referenced sample to evaluate detection at the county-year level. The three-signal composite achieves an out-of-sample area under the receiver operating characteristic curve of 0.87 on the validation sample, compared with 0.76 for nightlight alone, 0.78 for thermal anomalies alone, and 0.82 for rooftop classification alone. We report two final county-year measures. The first is a continuous activity index given by the composite score. The second is a confidence-weighted binary indicator that equals one when at least two of the three signals exceed their detection thresholds.

The geographic distribution of the resulting high-confidence measure is mapped in Appendix Figure B5.

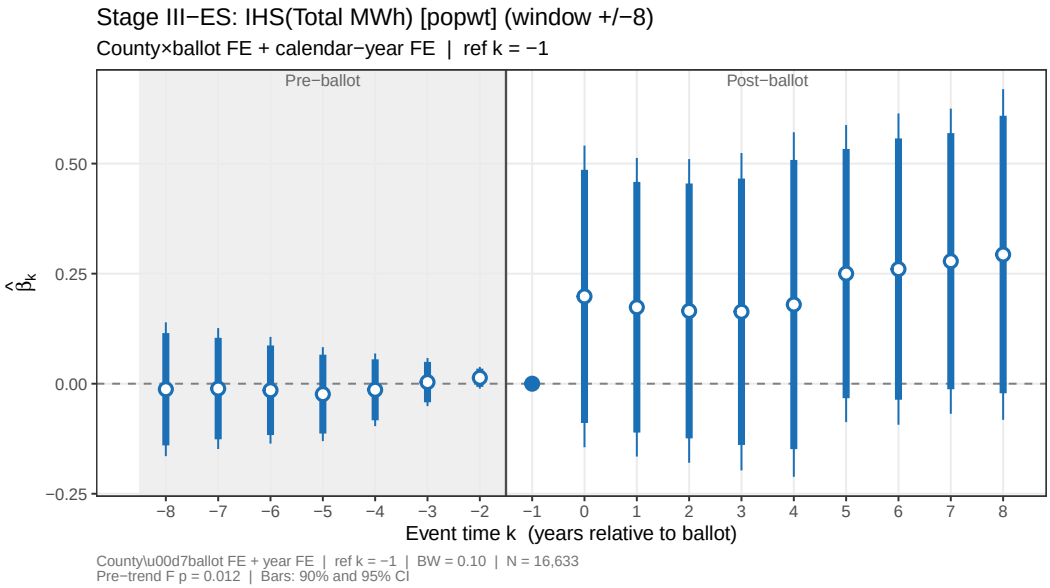
B Additional Figures

Figure B1. Density of the running variable at the cutoff



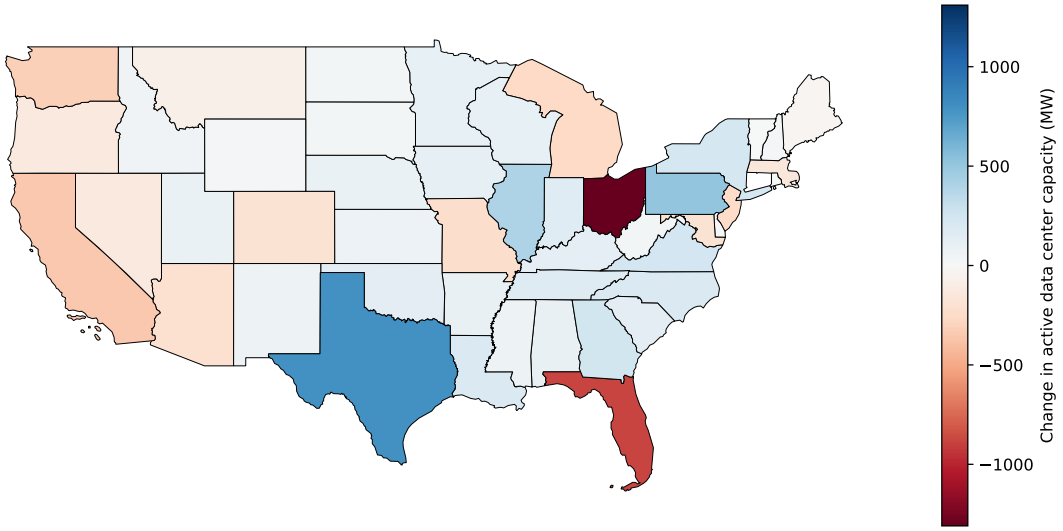
Notes: The figure plots the local-polynomial estimate of the density of the running variable R_{sa} on each side of the fifty-percent cutoff, with 95% confidence bands, following Cattaneo et al. (2020a). A discontinuity at the threshold would indicate that states sort across the pass-fail margin. The density is continuous at the cutoff on the twenty-six ballots within the bandwidth $|R_{sa}| \leq 0.10$ and on all thirty-four ballots. Source: authors' calculations.

Figure B2. Total electricity consumption around legalization at the eight-year window



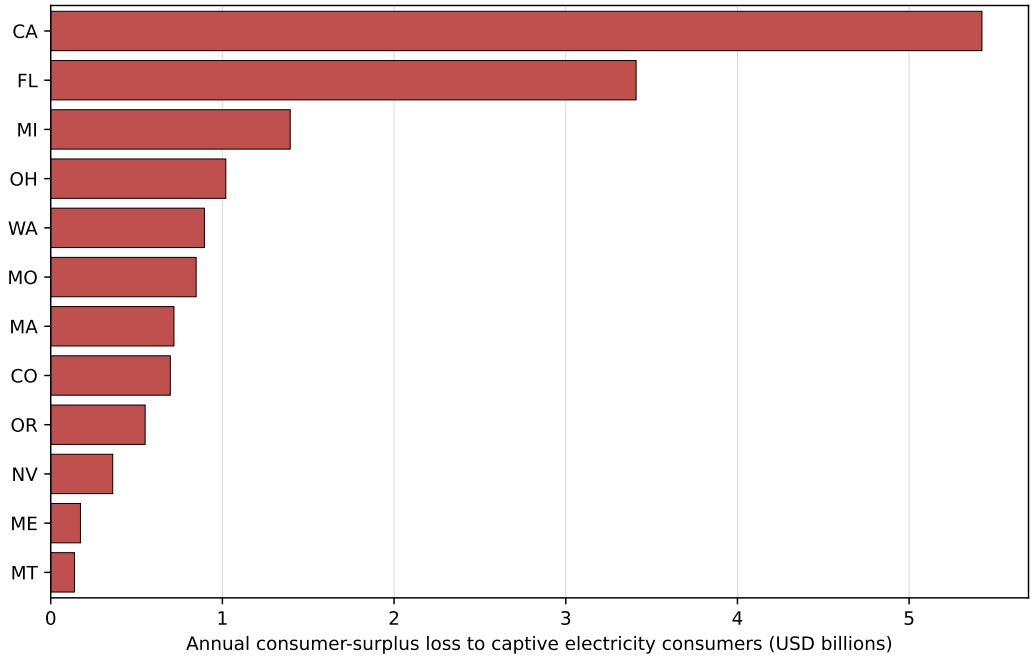
Notes: The figure reports a panel difference-in-discontinuities event study for population-weighted total electricity consumption in log megawatt-hours, estimated over a plus or minus eight-year window, and the year before the ballot is the omitted reference period. The 95% confidence intervals are based on standard errors clustered at the ballot level. The estimation sample is restricted to $|R_{sa}| \leq 0.10$. The plus or minus five-year version appears in the main text as Figure 7.

Figure B3. Counterfactual reallocation of U.S. data center capacity absent recreational cannabis legalization



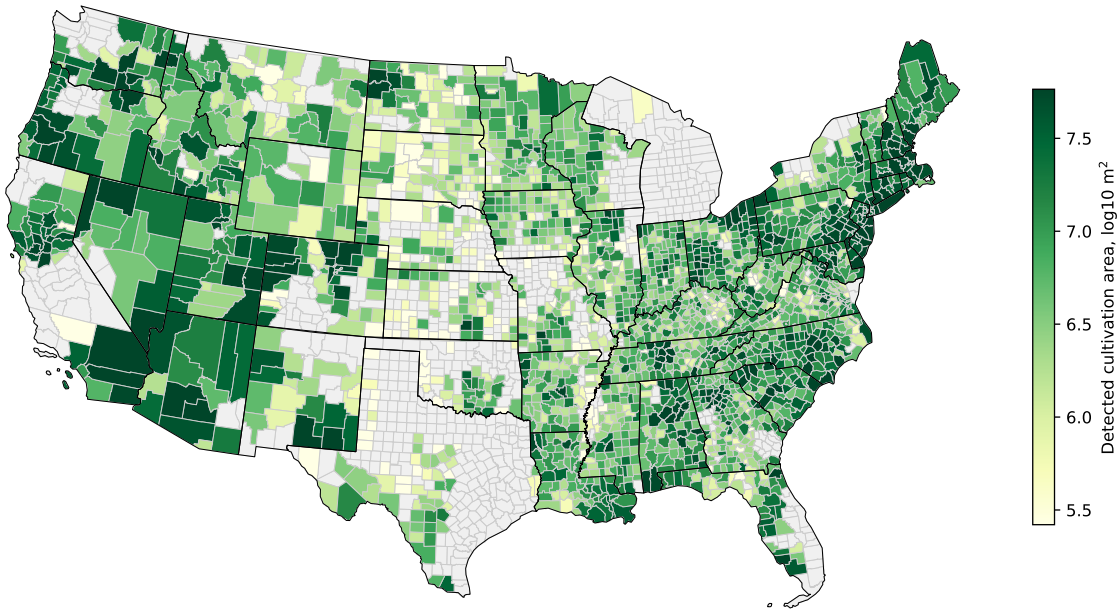
Notes: States are shaded by the 2024 data center capacity that the model attributes to legalization, measured as modeled capacity under observed prices minus modeled capacity under the counterfactual prices, holding the observed national total fixed. Blue shades mark states whose capacity is higher because of legalization elsewhere, the destinations of the reallocation, and red shades mark ballot-passing states whose capacity the price wedge pushes away. The counterfactual is computed by rolling back the legalization-induced electricity-price increase in treated counties and reallocating capacity across counties in proportion to the change in choice probabilities. Continental U.S. shown, with Alaska and Hawaii omitted from the map. Source: authors' calculations from the Aterio data center database and EIA Form 861.

Figure B4. State distribution of the annual consumer-surplus loss from recreational cannabis legalization



Notes: Each bar is a ballot-passing state's annual consumer-surplus loss to captive electricity consumers, the post-legalization retail-price wedge applied to 2024 county electricity consumption net of the mobile data-center share, summed over the state's counties inside the estimation bandwidth $|R_{sa}| \leq 0.10$. The bars sum to the 15.6 billion dollar bandwidth total reported in Section 6.3. States are ordered by loss. Source: authors' calculations from the location-choice model, EIA county electricity consumption, and the estimated first-stage legalization price effect.

Figure B5. County-level detected cannabis cultivation, 2024



Notes: Counties are shaded by the log of the detected high-confidence cultivation area for 2024, on a yellow-to-green sequential color scale. Counties with zero detected cultivation are shown in light gray. State outlines are overlaid in black. The detection measure is the three-signal composite described in Appendix A. Continental U.S. shown, with Alaska and Hawaii omitted from the map. Source: authors' calculations from VIIRS nightlight, Landsat thermal, and Sentinel-2 imagery.

C Additional Tables

Table C1. Cannabis Legalization and Data Center Stock: Three-Stage RD Summary

Outcome:	Stock		
$\hat{\beta}$ S-I	-1.355		
	(0.8572)		
$\hat{\beta}$ S-II		-1.464**	
		(0.7014)	
$\hat{\beta}$ S-III (Pass×Post)			-0.7019***
			(0.2194)
Fixed-effects			
county_fips-ballot_id			Yes
year			Yes
Observations	690	1,098	16,870
R^2	0.01339	0.01256	0.92833
Mean of dep. var.	0.462	0.436	0.446

Clustered (ballot_id) standard-errors in parentheses
*Signif. Codes: ***: 0.01, **: 0.05, *: 0.1*

Col (1): Stage I (single-ballot CS-RD, $k = 1$, no FE). Col (2): Stage II (pooled CS-RD, $k = 1$, no FE). Col (3): Stage III (panel RD-DiD, county×ballot FE + year FE). Level outcomes (raw count or MW). BW $|R| \leq 0.10$. SE clustered by ballot.

Table C2. Single-ballot cross-sectional estimates by data center outcome

Outcome:	Stock	Entry	StockMW	EntryMW	NoGen	Gen	Pipeline
$\hat{\beta}$: Pass	-1.355	-0.1084*	-19.97	-4.532	-0.7566	-0.5985*	-0.6321
	(0.8572)	(0.0556)	(12.38)	(3.642)	(0.6516)	(0.3193)	(0.8170)
Observations	690	690	690	690	690	690	690
R^2	0.01339	0.00695	0.00988	0.00586	0.01253	0.00879	0.00303
Mean of dep. var.	0.462	0.029	6.876	1.239	0.309	0.154	0.213

Clustered (ballot_id) standard-errors in parentheses
*Signif. Codes: ***: 0.01, **: 0.05, *: 0.1*

Single-ballot CS-RD at $k = 1$ (one year after ballot). No fixed effects. BW $|R_s| \leq 0.10$. SE clustered by ballot ($G = 13$).

Table C3. Pooled cross-sectional estimates by data center outcome

Outcome:	Stock	Entry	StockMW	EntryMW	NoGen	Gen	Pipeline
$\hat{\beta}$: Pass	-1.464**	-0.1109**	-18.97**	-2.545	-0.9570*	-0.5071**	-0.2812
	(0.7014)	(0.0529)	(8.854)	(1.561)	(0.5140)	(0.2123)	(0.3343)
Observations	1,098	1,098	1,098	1,098	1,098	1,098	1,098
R^2	0.01256	0.00555	0.01033	0.00723	0.01159	0.00946	0.00465
Mean of dep. var.	0.436	0.027	6.033	0.903	0.301	0.135	0.144

Clustered (ballot_id) standard-errors in parentheses
*Signif. Codes: ***: 0.01, **: 0.05, *: 0.1*

Pooled CS-RD at $k = 1$. No ballot-period FE (baseline specification). BW $|R_{sa}| \leq 0.10$. SE clustered by ballot ($G = 23$).

Table C4. Panel difference-in-discontinuities estimates by data center outcome, levels

Outcome:	Stock	Entry	StockMW	EntryMW	NoGen	Gen	Pipeline
$\hat{\beta}$: Pass×Post	-0.7019***	-0.0886*	-16.19**	-4.394**	-0.2567***	-0.4452**	-0.3656
	(0.2194)	(0.0447)	(7.620)	(2.063)	(0.0847)	(0.1626)	(0.2935)
Fixed-effects							
county_fips-ballot_id	Yes	Yes	Yes	Yes	Yes	Yes	Yes
year	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	16,870	16,870	16,870	16,870	16,870	16,870	16,870
R^2	0.92833	0.35427	0.70371	0.27831	0.96300	0.80055	0.29634
Mean of dep. var.	0.446	0.023	5.714	0.634	0.320	0.127	0.058

Clustered (ballot_id) standard-errors in parentheses
*Signif. Codes: ***: 0.01, **: 0.05, *: 0.1*

Pooled panel RD-DiD. County×ballot FE + calendar-year FE. (Year×ballot FE absorbs Pass × Post in single-state stacks; replaced with year FE to restore identification.) Controls: $R_{sa} \times$ Post and $R_{sa} \times$ Pass × Post. BW $|R_{sa}| \leq 0.10$, window ±8 yr. SE clustered by ballot ($G = 26$).

Table C5. Covariate balance: post-ballot change in regulatory environment (±8 yr, $|R_{sa}| \leq 0.10$)

Outcome:	RPS (binding)	RGGI member	Right-to-work	Retail dereg	Corp tax (top)
Pass×Post	-0.0160	0.0618*	-0.0157	-2.94×10^{-13}	-0.8835
	(0.0910)	(0.0351)	(0.0576)	(1.05×10^{-8})	(0.8341)
Fixed-effects					
county_fips-ballot_id	Yes	Yes	Yes	Yes	Yes
year	Yes	Yes	Yes	Yes	Yes
Fit statistics					
Observations	16,870	16,870	16,870	16,870	16,870

Clustered (ballot_id) standard-errors in parentheses
*Signif. Codes: ***: 0.01, **: 0.05, *: 0.1*

Each column reports the Pass×Post coefficient from the canonical RD-DiD applied to the state-year regulatory covariate listed in the header, merged onto the stacked county-year panel by ballot state. A cannabis ballot may signal voter willingness to adopt other regulation that also affects DC siting; if so, these covariates would jump at the cutoff after passage. None does. Covariate definitions: RPS = state has a binding mandatory Renewable Portfolio Standard in that year (DSIRE/NCSL); RGGI = state is a current participating member of the Regional Greenhouse Gas Initiative; Right-to-work = state has a right-to-work statute (NCSL); Retail dereg = state has full retail electricity choice (EIA); Corp tax = top statutory state corporate income tax rate, in percent (ITEP Historical State Tax Rate Data, compiled from FTA). County×ballot and year FE; BW $|R_{sa}| \leq 0.10$; SE clustered by ballot. $^{*}p < .1, ^{**}p < .05, ^{***}p < .01$.

Table C6. Stage III robustness: not-yet-treated control exclusion

Outcome	Active stock		Active entry	
Specification	Baseline	Excl. legalized	Baseline	Excl. legalized
$\hat{\beta}$: Pass×Post	-0.7019***	-0.4582**	-0.0886*	-0.0357
	(0.2194)	(0.1654)	(0.0447)	(0.0242)
Fixed-effects				
county_fips-ballot_id	Yes	Yes	Yes	Yes
year	Yes	Yes	Yes	Yes
Observations	16,870	16,369	16,870	16,369
R^2	0.92833	0.94347	0.35427	0.38702
Mean of dep. var.	0.446	0.413	0.023	0.020

Clustered (ballot_id) standard-errors in parentheses
*Signif. Codes: ***: 0.01, **: 0.05, *: 0.1*

Stage III panel difference-in-discontinuities (county×ballot FE + calendar-year FE; controls $R_{sa} \times \text{Post}$ and $R_{sa} \times \text{Pass} \times \text{Post}$; BW $|R_{sa}| \leq 0.10$, window ±8 yr). Excl. legalized drops each control state's observations from the year the state itself legalizes onward (California from 2016, Oregon from 2014, Arizona from 2020); South Dakota's overturned 2020 initiative and Florida's failed 60 percent supermajority are not treated as legalization; treating South Dakota's 2020 majority as legalization instead gives −0.556 (stock) and −0.046 (entry). At a three-year window the baseline and excluded stock samples give −0.196 and −0.192; the excluded states supply three of the six control ballots at the seventh and eighth post-ballot years. SE clustered by ballot ($G = 26$).

Table C7. The data center stock effect across event windows

	±3 yr	±6 yr	±8 yr
$\hat{\beta}$: Pass×Post	-0.1961**	-0.4963***	-0.7019***
	(0.07175)	(0.1585)	(0.2194)
Fixed-effects			
County × ballot	Yes	Yes	Yes
Calendar year	Yes	Yes	Yes
Observations	7,930	13,538	16,870
R^2	0.98131	0.95786	0.92833
Mean of dep. var.	0.402	0.427	0.446

Clustered (ballot_id) standard-errors in parentheses
*Signif. Codes: ***: 0.01, **: 0.05, *: 0.1*

The panel difference-in-discontinuities specification estimates the Pass×Post coefficient on the active data center count at three event windows, all with county-by-ballot and calendar-year fixed effects and estimation bandwidth $|R_{sa}| \leq 0.10$. The magnitude grows with the window because the pass-fail gap accumulates across post-ballot years, consistent with a permanent rather than a transient reduction.

Table C8. The data center stock and entry effects across bandwidths

Bandwidth $\lvert R_{sa}\rvert$	G_-	G_+	$\hat{\beta}$ Stock	$\hat{\beta}$ Entry
0.04	5	4	-1.193***	-0.244***
0.05	6	6	-1.035**	-0.177*
0.06	9	10	-0.787**	-0.117
0.07	10	12	-0.741**	-0.113*
0.08	11	14	-0.839***	-0.142**
0.10	12	14	-0.702***	-0.089*
0.12	13	15	-0.722***	-0.114**
0.15	14	15	-0.642***	-0.088*

Clustered (ballot_id) standard-errors in parentheses
*Signif. Codes: ***: 0.01, **: 0.05, *: 0.1*

The panel difference-in-discontinuities specification estimates the Pass×Post coefficient on the active stock and on new entries across bandwidths. G_- and G_+ are the ballot events below and above the cutoff within each bandwidth. The event window is plus or minus eight years around the ballot, and all specifications include county-by-ballot and calendar-year fixed effects. The estimate is stable from 0.07 to 0.15 and strengthens at the narrowest bandwidths, reaching 1.19 fewer facilities at the mean-squared-error optimal bandwidth of 0.04. Because that bandwidth leaves only four or five ballots per side, too few for reliable cluster-robust inference, the baseline holds the bandwidth at 0.10.

Table C9. The data center stock effect across fixed-effect structures

	County	County×ballot	County×ballot + year
$\hat{\beta}$: Pass×Post	-0.6839**	-0.7188***	-0.7019***
	(0.2667)	(0.2485)	(0.2194)
Fixed-effects			
County	Yes		
County × ballot		Yes	Yes
Calendar year			Yes
Observations	16,870	16,870	16,870
R^2	0.90647	0.92740	0.92833
Mean of dep. var.	0.446	0.446	0.446

Clustered (ballot_id) standard-errors in parentheses
*Signif. Codes: ***: 0.01, **: 0.05, *: 0.1*

The table reports the panel difference-in-discontinuities Pass×Post coefficient on the active data center stock under three fixed-effect structures, each including the ballot-specific post indicator. The event window is ±8 years around the ballot and the estimation bandwidth is $|R_{sa}| \leq 0.10$. The estimate is stable near -0.70 across every fixed-effect structure, confirming the contrast reflects the treatment effect rather than a fixed-effects artifact.

Table C10. Stage III robustness: running-variable polynomial order

Outcome	Active stock		Active entry	
Specification	Baseline	Quadratic R_{sa}	Baseline	Quadratic R_{sa}
$\hat{\beta}$: Pass×Post	-0.7019***	-0.9930**	-0.0886*	-0.2000**
	(0.2194)	(0.3962)	(0.0447)	(0.0862)
Fixed-effects				
county_fips-ballot_id	Yes	Yes	Yes	Yes
year	Yes	Yes	Yes	Yes
Observations	16,870	16,870	16,870	16,870
R^2	0.92833	0.92846	0.35427	0.35478
Mean of dep. var.	0.446	0.446	0.023	0.023

Clustered (ballot_id) standard-errors in parentheses
*Signif. Codes: ***: 0.01, **: 0.05, *: 0.1*

Stage III panel difference-in-discontinuities (county×ballot FE + calendar-year FE; controls $R_{sa} \times$ Post and $R_{sa} \times$ Pass × Post; BW $|R_{sa}| \leq 0.10$, window ±8 yr). Quadratic R_{sa} adds $R_{sa}^2 \times$ Post and $R_{sa}^2 \times$ Pass × Post to the running-variable polynomial; R_{sa} and R_{sa}^2 are ballot-constant and absorbed by the county-by-ballot fixed effects. SE clustered by ballot ($G = 26$).

Table C11. Wild cluster bootstrap inference

Outcome	$\hat{\beta}$	G_-/G_+	p analytical	p bootstrap
Active stock	-0.702	12/14	0.004	0.028
Active entry	-0.089	12/14	0.059	0.147

Standard errors clustered by ballot_id

The panel difference-in-discontinuities specification is estimated at the bandwidth $|R_{sa}| \leq 0.10$ over an event window of plus or minus eight years around the ballot. Inference uses the studentized wild cluster bootstrap with ballot-level Rademacher weights following Cameron et al. (2008). The active-stock effect remains significant at the 5 percent level under the bootstrap.

Table C12. Cannabis legalization and detected cultivation

	(1)	(2)	(3)	(4)	(5)	(6)
$\hat{\beta}$ single-ballot cross-section	-4.342	-0.2403				
	(2.971)	(0.2496)				
$\hat{\beta}$ pooled cross-section			0.8835	0.1157		
			(1.284)	(0.1140)		
$\hat{\beta}$ panel diff-in-disc (Pass×Post)					0.4879	0.0808
					(0.6081)	(0.0789)
Fixed effects						
County × ballot					Yes	Yes
Calendar year					Yes	Yes
Observations	690	690	1,098	1,098	16,870	16,870
R^2	0.04656	0.01878	0.03468	0.01149	0.15179	0.10008
Within R^2					0.00173	0.00022

Clustered (ballot_id) standard-errors in parentheses
*Signif. Codes: ***: 0.01, **: 0.05, *: 0.1*

The table reports estimates of the effect of recreational cannabis legalization on detected cultivation, at the county-year level. Columns (1), (3), and (5) use the inverse hyperbolic sine $\sinh^{-1}(y)$ as the dependent variable, and Columns (2), (4), and (6) use the cultivation area in km^2 scaled by 10^{-6}. The single-ballot and pooled cross-sections are estimated by cross-sectional regression discontinuity at $k = 1$ with no fixed effects, while the panel difference-in-discontinuities specification includes county-by-ballot and calendar-year fixed effects. The estimation bandwidth is $|R| \leq 0.10$.

Table C13. Electricity price effect: full sample, detection-window sample, and cultivation-intensity interaction

	Average			Commercial			Industrial		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
$\hat{\beta}_1$: Pass×Post	0.1243	0.1199	0.1197	0.1178	0.1823	0.1823	0.1035	0.1449	0.1444
	(0.1414)	(0.1064)	(0.1051)	(0.1282)	(0.1341)	(0.1331)	(0.1337)	(0.1195)	(0.1179)
$\hat{\beta}_2$: km^2×Pass×Post			0.0219***			0.0231***			0.0264*
			(0.0072)			(0.0066)			(0.0136)
Observations	11,563	6,868	6,868	11,563	6,868	6,868	11,355	6,841	6,841
Sample	Full	Restricted	Restricted	Full	Restricted	Restricted	Full	Restricted	Restricted
km^2 interaction	No	No	Yes	No	No	Yes	No	No	Yes

Notes: Stage III panel RD-DiD on log electricity prices, county-year level. County×ballot FE + calendar-year FE. BW $|R_{sa}| \leq 0.10$. SE clustered by ballot. Window ±5 yr. "Full" uses all ballot-state counties in the bandwidth; Restricted sample = 2017–2024 (satellite 3-ind coverage). The restricted sample retains 19 of the 26 ballots: four have no post-2017 window, and three retain a single calendar year and are absorbed by the county-by-ballot fixed effects. Columns (3), (6) and (9) add $\hat{\beta}_2$, the marginal effect per km^2 of detected cannabis in a legalized county post-ballot, which is defined only on the restricted sample. Stars: *** 0.01, ** 0.05, * 0.10.

Table C14. Electricity prices and detected cultivation in levels

	Average (1)	Commercial (2)	Industrial (3)
$\hat{\gamma}$: triconf area (km^2)	0.0035*	0.0019	0.0064***
	(0.0017)	(0.0021)	(0.0018)
Fixed effects			
County × ballot	Yes	Yes	Yes
Calendar year	Yes	Yes	Yes
Fit statistics			
Observations	6,868	6,868	6,841
R^2	0.84657	0.85500	0.89337
Within R^2	0.00072	0.00021	0.00112

Notes: The table reports within-county estimates of the association between detected cultivation and electricity prices, at the county-year level. The outcome in every column is the natural log of the population-weighted retail electricity price in dollars per kWh for a customer class, namely the all-customer average in Column (1), commercial in Column (2), and industrial in Column (3). The regressor is the level in km^2 of triple-confirmed satellite cultivation area. All specifications include county-by-ballot and calendar-year fixed effects. The estimation bandwidth is $|R_{sa}| \leq 0.10$, and the sample is restricted to 2017-2024. Standard errors are clustered at the ballot level, with 26 or fewer clusters, and are reported in parentheses. Three, two, and one asterisks denote statistical significance at the 1, 5, and 10 percent levels, respectively.

Table C15. Electricity prices and detected cultivation within county at the eight-year window

	Average (1)	Commercial (2)	Industrial (3)
$\hat{\gamma}$: inverse hyperbolic sine(triconf km^2)	0.0510***	0.0477***	0.0670***
	(0.0092)	(0.0132)	(0.0077)
Fixed effects			
County × ballot	Yes	Yes	Yes
Calendar year	Yes	Yes	Yes
Fit statistics			
Observations	8,649	8,649	8,591
R^2	0.86565	0.86644	0.89844
Within R^2	0.00606	0.00503	0.00526

Notes: The table reports within-county estimates of the association between detected cultivation and electricity prices, at the county-year level. The outcome in every column is the natural log of the population-weighted retail electricity price in dollars per kWh for a customer class, namely the all-customer average in Column (1), commercial in Column (2), and industrial in Column (3). The regressor is the inverse hyperbolic sine of triple-confirmed satellite cultivation area. All specifications include county-by-ballot and calendar-year fixed effects. The estimation bandwidth is $|R_{sa}| \leq 0.10$, and the sample is restricted to 2017-2024. Standard errors are clustered at the ballot level, with 26 or fewer clusters, and are reported in parentheses. Three, two, and one asterisks denote statistical significance at the 1, 5, and 10 percent levels, respectively.

Table C16. Cannabis cultivation, electricity prices, and consumption at the eight-year window

	Total		Commercial		Industrial	
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A. Electricity price						
$\hat{\beta}_1$: Pass×Post	0.1213	0.1209	0.1817	0.1816	0.1435	0.1427
	(0.1027)	(0.1020)	(0.1427)	(0.1423)	(0.1120)	(0.1112)
$\hat{\beta}_2$: km^2×Pass×Post		0.0110*		0.0146*		0.0116
		(0.0064)		(0.0073)		(0.0082)
Observations	8,649	8,649	8,649	8,649	8,591	8,591
Panel B. Electricity consumption						
$\hat{\beta}_1$: Pass×Post	0.0321	0.0325	-0.0647	-0.0645	-0.0046	-0.0042
	(0.3657)	(0.3655)	(0.3963)	(0.3962)	(0.2075)	(0.2074)
$\hat{\beta}_2$: km^2×Pass×Post		0.0067		0.0045		0.0015
		(0.0105)		(0.0138)		(0.0124)
Observations	8,649	8,649	8,649	8,649	8,649	8,649
km^2 interaction	No	Yes	No	Yes	No	Yes

Notes: Stage III panel RD-DiD, estimated exactly as in Table 2 but on the eight-year event window. County×ballot FE + calendar-year FE. BW $|R_{sa}| \leq 0.10$. SE clustered by ballot. Window ±8 yr. Both panels use the cannabis-restricted sample (2017–2024 satellite 3-ind coverage), so the two panels are estimated on the same county-years and the interaction columns are defined throughout. "Total" is the average retail price in Panel A and total consumption in Panel B. Price outcomes are in logs; consumption outcomes in inverse hyperbolic sine of MWh. Odd columns omit the interaction; even columns add $\hat{\beta}_2$, the marginal effect per km^2 of detected cannabis in a legalized county post-ballot. Stars: *** 0.01, ** 0.05, * 0.10.

Table C17. Industry employment and share of county labor

Industry (NAICS)	Mean emp. per county-year (workers)	Mean industry share of county total emp. (%)
Cannabis cultivation, indoor (NAICS 111419)	2.6	0.003%
Data centers (NAICS 518210)	110.9	0.042%
Information (NAICS 51)	868.1	0.784%
Construction (NAICS 23)	2,239.8	4.62%
Manufacturing (NAICS 31-33)	3,820.3	11.31%
Accommodation & food services (NAICS 72)	3,991.5	6.98%
Retail trade (NAICS 44-45)	4,754.2	11.43%
County total (NAICS 10)	44,974.2	100.00%

Notes: All values are per-year, per-county figures from BLS QCEW, averaged across 22,914 county-year cells, which is 7 years times 3,275 counties. Column 1 reports mean employment per county-year in workers, which is the mean across county-year cells of `annual_avg_emplvl`, the BLS annual average of monthly employment headcount for that county and NAICS in that year. The unit is workers rather than worker-years summed across the sample. County-years with no firm in the relevant NAICS contribute zero. The county-total row averages 44,974 workers, which is the mean across counties of a typical year's total employment. This figure is heavily pulled upward by the largest urban counties such as LA County, Cook County, and Harris County, each of which employs millions, and the median county is far smaller. The aggregate is internally consistent with national US employment, since 44,974 times 3,275 counties is about 147 million workers per year, matching the BLS published US covered employment totals for 2017-2023. Column 2 reports mean industry share of county total employment, which is computed for each county-year as industry employment divided by county total employment and then averaged across county-years. This is the share of the local labor market employed in each industry in a typical county-year. Because the share is computed at the county-year level and then averaged, it is not affected by the urban-county skew that inflates the county-total mean. Cannabis indoor and data centers are both well under 1% of county employment, while traditional labor-intensive sectors such as manufacturing, accommodation and food, and retail are 5-12%.

Table C18. Top occupations in data centers and cannabis cultivation

	Data centers NAICS 518200		Support Activities for Crop Production NAICS 115100		Agriculture, Forestry, Fishing and Hunting sector NAICS 11	
Rank	Occupation (SOC)	Share	Occupation (SOC)	Share	Occupation (SOC)	Share
1	Software Developers (15-1252)	16.1%	Farmworkers and Laborers, Crop, Nursery, and Greenhouse (45-2092)	55.4%	Farmworkers and Laborers, Crop, Nursery, and Greenhouse (45-2092)	45.7%
2	Sales Reps. of Services, Except Advertising, Insurance, Financial Services, and Travel (41-3091)	6.3%	Packers and Packagers, Hand (53-7064)	6.3%	Packers and Packagers, Hand (53-7064)	5.2%
3	Customer Service Representatives (43-4051)	5.4%	Agricultural Equipment Operators (45-2091)	4.5%	Logging Equipment Operators (45-4022)	4.0%
4	Computer and Information Systems Managers (11-3021)	5.0%	First-Line Supervisors of Farming, Fishing, and Forestry Workers (45-1011)	3.2%	Agricultural Equipment Operators (45-2091)	3.8%
5	Computer User Support Specialists (15-1232)	3.9%	Graders and Sorters, Agricultural Products (45-2041)	3.0%	First-Line Supervisors of Farming, Fishing, and Forestry Workers (45-1011)	3.6%
6	General and Operations Managers (11-1021)	3.9%	Laborers and Freight, Stock, and Material Movers, Hand (53-7062)	2.8%	Heavy and Tractor-Trailer Truck Drivers (53-3032)	3.2%

Continued on next page

(Table C18 continued)

Rank	Occupation (SOC)	Share	Occupation (SOC)	Share	Occupation (SOC)	Share
7	Market Research Analysts and Marketing Specialists (13-1161)	3.2%	Industrial Truck and Tractor Operators (53-7051)	2.3%	Farmworkers, Farm, Ranch, and Aquacultural Animals (45-2093)	3.0%
8	Computer Systems Analysts (15-1211)	2.8%	Heavy and Tractor-Trailer Truck Drivers (53-3032)	1.6%	Graders and Sorters, Agricultural Products (45-2041)	2.5%
9	Computer Occupations, All Other (15-1299)	2.8%	Packaging and Filling Machine Operators and Tenders (51-9111)	1.6%	Laborers and Freight, Stock, and Material Movers, Hand (53-7062)	2.4%
10	Project Management Specialists (13-1082)	2.5%	Office Clerks, General (43-9061)	1.3%	Industrial Truck and Tractor Operators (53-7051)	1.9%
11	Office Clerks, General (43-9061)	1.9%	General and Operations Managers (11-1021)	1.2%	Office Clerks, General (43-9061)	1.7%
12	Software QA Analysts and Testers (15-1253)	1.9%	Farm Equipment Mechanics and Service Technicians (49-3041)	0.9%	Animal Caretakers (39-2021)	1.5%
	Total emp: 484,990		*Total emp: 351,890*		*Total emp: 427,250*	

Notes: The table reports the top 12 detailed-SOC occupations within each industry, ranked by employment, from the BLS OEWS May 2024 national release. "Share" is the percentage of the industry's total employment in that occupation, taken from the OEWS `PCT_TOTAL` field. NAICS 115100, Support Activities for Crop Production, is the closest OEWS-published proxy for cannabis-cultivation labor. The direct crop-production codes NAICS 1114 and 1119, which contain the cannabis cultivation codes 111419 and 111998, are not published in OEWS. The NAICS 11 sector total in OEWS comprises only logging and support activities, about 427,000 workers, and the direct agriculture production codes are not included in the sector total either.

Table C19. Weekly wage across broad sectors

Sector	$\hat{\beta}$: Pass×Post ($/wk)	Pre-period mean ($/wk)	Observations	Within R^2
Construction	-61.24*	860.0	14,850	0.009
	(33.26)			
Manufacturing	-66.22	891.1	13,760	0.004
	(46.16)			
Retail trade	-8.51	490.2	16,276	0.016
	(20.45)			
Information	-261.34**	886.4	12,678	0.008
	(111.03)			
Accommodation & food	5.53	295.1	13,109	0.035
	(23.99)			
Leisure & hospitality	9.14	294.5	16,380	0.045
	(26.15)			
Professional services	-51.33	850.1	15,445	0.002
	(70.10)			
County × ballot FE	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes

Notes: The table reports panel difference-in-discontinuities estimates of the effect of recreational cannabis legalization on the intensive-margin weekly wage in seven broad sectors, at the county-year level, dropping county-years coded NA. All specifications include county-by-ballot and calendar-year fixed effects, and the estimation bandwidth is $|R_{sa}| \leq 0.10$ with an estimation window of eight years either side of the ballot. Standard errors are clustered at the ballot level and are reported in parentheses. No sector shows the positive wage response that labor competition requires, and the significant coefficients are negative, consistent with demand-side adjustment rather than competition for a shared labor pool. Three, two, and one asterisks denote statistical significance at the 1, 5, and 10 percent levels, respectively.

Table C20. Cannabis cultivation wage by indoor and outdoor segment

	Agg. 6-digit (in.+out.)	Indoor only 111419	Outdoor only 111998	Agg. 5-digit (in.+out.)	Indoor only 11141	Outdoor only 11199
	(1)	(2)	(3)	(4)	(5)	(6)
Pass × Post	95.6	619.0**	11.0	67.6	159.6	22.0
	(174.4)	(256.1)	(115.9)	(123.1)	(173.3)	(106.6)
Pre-period mean ($/wk)	553.5	549.6	552.5	575.0	550.7	570.8
County × Ballot FE	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes
Observations	2,307	562	1,998	2,691	978	2,263
R^2	0.673	0.834	0.643	0.674	0.783	0.650
Within R^2	0.002	0.056	0.001	0.003	0.009	0.001

Notes: The table reports pooled panel difference-in-discontinuities estimates of the effect of recreational cannabis legalization on cannabis-cultivation wages, at the county-year level. The outcome is the weekly wage in dollars per worker. County-years coded NA are dropped, so the sample is restricted to county-years where at least one firm exists in the relevant NAICS aggregate and BLS disclosed the cell. Columns (1) and (4) are employment-weighted averages across the indoor and outdoor component NAICS codes, while Columns (2), (3), (5), and (6) are single-NAICS measures. All specifications include county-by-ballot and calendar-year fixed effects, the estimation bandwidth is $|R_{sa}| \leq 0.10$, and the event window is plus or minus eight years around the ballot. The 5-digit indoor code NAICS 11141 is the residual not assigned to a 6-digit subcode, and the 5-digit outdoor code NAICS 11199 is the aggregate of other crop farming including outdoor cannabis. Standard errors are clustered at the ballot level and are reported in parentheses. Three, two, and one asterisks denote statistical significance at the 1, 5, and 10 percent levels, respectively.

Table C21. Cannabis legalization and state trade in AI-related products

Triple difference	Coefficient	SE	Ballots
Imports			
All AI-related	-0.034	(0.089)	18
Compute hardware	-0.433*	(0.237)	18
Servers (HS 8471)	-0.354**	(0.158)	18
Exports			
All AI-related	0.061	(0.107)	18
Compute hardware	-0.129	(0.210)	18
Servers (HS 8471)	-0.245	(0.154)	18

Clustered (ballot) standard errors in parentheses
*Signif. Codes: ***: 0.01, **: 0.05, *: 0.1*

The outcome is the inverse hyperbolic sine of a state's annual trade. Each cell is the triple-difference coefficient on $Pass \times Post \times AI$, comparing trade in AI-related products against all other trade within the balloting state. *All AI-related* uses every product classified as High relevance to AI data-center infrastructure by Waugh (2026); *Compute hardware* restricts to the GPUs, servers, and storage of that classification's Compute Hardware category, the imported, data-center-specific equipment. State trade is published at HS6 and the classification is at HS10, so each HS6 heading is split by the national share of its value in the relevant HS10 lines that year. The Servers row narrows the compute-hardware measure to HS heading 8471, automatic data processing machines, the single most data-center-specific import. Fixed effects are ballot-by-AI, year-by-AI, and ballot-by-year. The compute-hardware and servers import estimates have wild cluster bootstrap p-values of 0.079 and 0.036. Coverage: 2013-2024, 18 ballots inside the bandwidth with pre- and post-ballot trade. Unit of analysis: state-year. Bandwidth $|R_s| \leq 0.10$. Standard errors clustered by ballot.