Grid Lock-in and the Development of AI*

Xinming Du† Jingyuan Guo‡

September 27, 2026

Abstract

When electricity infrastructure adjusts slowly, expansion in one electricity-intensive industry can constrain subsequent local investment in another. We study cannabis cultivation and data centers using close recreational cannabis ballots and a new satellite measure of indoor cultivation. Legalization raises detected cultivation by 3.6 times the sample mean, implying 22 megawatts of demand per county. The data center stock falls by 0.70 facilities and 16.2 megawatts per county, and retail electricity prices rise where cultivation expands. A siting model implies that, absent legalization, annual electricity bills in narrowly passing states would be over 12 billion dollars lower.

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

*We thank Mark Colas, David Lagakos, Chen Liu, and Wolfram Schlenker for helpful comments. We used AI-assisted tools for editing in accordance with journal guidelines. All analysis, arguments, and conclusions are our own.

†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

Can expanding one industry today limit a region's ability to attract another industry tomorrow? New industrial activity may benefit other local industries, through local demand and productivity spillovers (Greenstone et al., 2010; Moretti, 2010). However, industries could compete for local inputs that are fixed in the short run, such as housing (Hsieh and Moretti, 2019) and workers (Allcott and Keniston, 2018). When such an input is scarce and slow to expand, the order in which industries arrive can affect whether later investment takes place there, and an early arrival can shape a region's path long after it occurs (David, 1985; Arthur, 1989).

We study this question in the U.S. electricity grid. Generation and transmission take years to build (Rand et al., 2024), while large new loads can arrive much faster. Transmission construction has slowed sharply over the past decade. Only 322 miles of transmission at 345 kilovolts and above were completed in 2024, against nearly 4,000 miles in 2013 (Shreve et al., 2025). If one industry absorbs local grid capacity, later electricity-intensive firms may face higher power costs or longer delays in connecting to the grid, which can deter their investment in the region. We refer to this constraint on local development as grid lock-in.

This constraint has become increasingly relevant for data centers, the physical infrastructure of AI. They consume 4-5 percent of U.S. electricity today and could reach 9-17 percent by 2030 (Electric Power Research Institute, 2026). A new facility costs 9.5-13.3 million dollars per megawatt before any computing hardware (Turner & Townsend, 2025), so the siting of a new facility represents a large local investment. Local grid conditions can therefore affect whether new capacity is built in a region as well as the cost of operating it (Kahn and Mansur, 2013).

Identifying grid lock-in is difficult, because grid conditions and industrial investment are jointly determined. Places that attract one large electricity user may also differ in ways that affect later entry. Observing no subsequent investment is also uninformative by itself, because the missing facility may never have been headed there. An answer therefore requires a shock to local electricity demand and a credible comparison group for how much later investment a region would otherwise have attracted.

Recreational cannabis legalization provides both the early electricity-intensive industry and the variation this question requires. Indoor cultivation 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). Close legalization ballots generate plausibly exogenous variation in this localized electricity load, and indoor cultivation shares little labor, technology, or product-market exposure with data centers. We compare counties in states where a ballot barely passed with

counties in states where it barely failed.

We assemble a national county-year panel spanning 2000-2024 that links data centers, electricity markets, cannabis ballots, and indoor cultivation. 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 consistent administrative 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 from 2017 onward (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 the estimate compares the change after narrowly passing ballots with the change after narrowly failing ones. A pre-existing outcome gap at the threshold is allowed, so long as the gap would not have changed had the ballot failed. Sorting across the threshold is a separate concern, and the density of vote shares shows no discontinuity at the majority threshold. The design also requires that a county's outcomes not depend on ballot outcomes in other states; dropping comparison counties served by a utility that also serves a passing state leaves the estimate essentially unchanged. The sample covers ballots within ten percentage points of the threshold, and we cluster by ballot.

In counties in 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 in 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. At U.S. construction costs, the lost capacity represents over 150 million dollars of forgone investment per county.

Detected indoor cultivation rises by 3.6 times the sample mean within five post-ballot years. The increase is 0.096 square kilometers per county in the ballot year and 0.176 five years after it, against a mean of 0.0485. Every detected pixel is confirmed by the rooftop

signal, so the increase is in enclosed, electricity-intensive cultivation. At the energy intensity of indoor cultivation, the five-year increase implies an average electricity demand of 22 MW per county, comparable to the 16.2 MW of data center capacity lost. In legalizing counties, retail electricity prices rise 2.2 percent more after the ballot for each square kilometer of detected cultivation. Total electricity consumption, which we observe only as utility sales allocated across the counties each utility serves, does not move measurably.

Data centers appear to respond not only to the capacity absorbed by cultivation, but also to the expected pressure on the shared grid. The decline starts in the ballot year, before cultivation has scaled. When counties are split by pre-ballot cannabis supply, the price increase is larger in the high-supply group, while data centers decline in both groups, and by more in the high-supply group. Data center activity also declines in counties in other states that share the grid, where the ballot does not legalize cultivation.

Competition for workers or land does not appear to drive the decline. Local wages do not rise. The estimated change in the county weekly wage is -6.2 dollars against a pre-period mean of 745. The two industries employ largely different occupations, software developers in data centers and farmworkers in cultivation. Land is a small share of a data center's capital outlay, and the two industries seek different sites: data centers need parcels near high-voltage substations and fiber, while cultivation occupies warehouse space.

We use a discrete-choice model of county-level data center siting to quantify the investment and electricity-cost effects. In the model, legalization lowers data center investment in the states that narrowly passed a ballot, and states that never held a recreational ballot, such as Texas, Pennsylvania, and Illinois, attract more of it. Annual electricity bills in the narrowly passing states would be over 12 billion dollars lower without legalization.

Our results make three contributions. First, we add to work on how the entry of one industry affects the local development of others. Existing work emphasizes 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; Black et al., 2005), as well as knowledge spillovers and product-market competition (Bloom et al., 2013; Moretti, 2021). We document a channel operating through shared electricity infrastructure. When grid capacity adjusts slowly, demand from one electricity-intensive industry can deter later entrants from investing in the region. The closest evidence is on cryptocurrency mining, whose electricity use raises the bills of existing households and small businesses where miners enter (Benetton et al., 2023). This channel connects to work on how transportation infrastructure (Donaldson and Hornbeck, 2016; Donaldson, 2018; Faber, 2014; Asher and Novosad, 2020) and housing supply constraints (Hsieh and Moretti, 2019) shape local economic development, and to the broader literature on lock-in and path dependence

(David, 1985; Arthur, 1989; Unruh, 2000).

Second, we contribute to work on investment in data centers and their local effects. Existing work documents their energy and water demands (Masanet et al., 2020; Siddik et al., 2021; Mytton, 2021; Shehabi et al., 2024) and their implications for electricity grids and ratepayers (International Energy Agency, 2024; Electric Power Research Institute, 2024), while recent studies examine their local economic effects and the determinants of their siting (Alvarez et al., 2026; Ancona et al., 2026; Du et al., 2026; Zhang, 2026). Our design uses close cannabis ballots to generate exogenous variation in electricity demand from another industry, which allows us to study whether competition for local grid capacity deters subsequent data center investment. Our results also relate to work on digital technology and AI, which has focused primarily on labor markets, productivity, and market structure and less on the physical infrastructure that computing requires (Acemoglu and Restrepo, 2018, 2020; Brynjolfsson et al., 2021; Goldfarb and Tucker, 2019).

Third, we identify an electricity-market spillover of recreational cannabis legalization. Existing work has studied the effects of cannabis laws on traffic safety and substance use (Anderson et al., 2013; Wen et al., 2015), crime (Adda et al., 2014; Gavrilova et al., 2019), and housing markets (Cheng et al., 2018). We show that legalization can also affect investment in an unrelated industry through competition for a shared local input. The decline in data centers also reaches counties on the same grid in states where the ballot did not legalize cultivation. We construct a satellite-based measure of indoor cultivation that is comparable across states with different reporting regimes (Henderson et al., 2012; Donaldson and Storeygard, 2016; Colas and Reynier, 2026).

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 & 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. 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, centered on 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. Electricity dominates their operating cost, and industrial retail rates vary by a factor of two or more across U.S. counties. A one-cent change in the price per kilowatt-hour changes the lifetime electricity cost of a 100-megawatt facility by 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. Plants receive 12-18 hours of light a day depending on growth stage, making electricity a first-order variable operating cost. Mills (2012) estimates that indoor cultivation uses eight times as much energy per square foot as a typical commercial building. Electricity use by cannabis cultivation and infused-product manufacturing grew from 1 percent of Denver's total in 2013 to 4 percent in 2018 (Denver Department of Public Health and Environment, 2019). Environmental control and lighting dominate the load, and the energy needed for an artificial indoor climate makes cannabis more emissions-intensive than conventional crop production (Summers et al., 2021).

Outdoor cultivation uses less electricity, but regulation and product-market incentives concentrate licensed production indoors (Mills, 2025). Security, video-surveillance, and seed-to-sale tracking requirements are more easily met in enclosed facilities. 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 predominantly indoor. Recreational legalization expands 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 most 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 legal 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 this expansion is the shock we study. Stores typically open within two years of the ballot. Most recreational adoptions have come through citizen ballot initiatives, a smaller number through state legislation. For the empirical design, ballot timing is determined well before election day through signature collection and certification, and the statewide vote decides passage, so contemporaneous conditions in any one county do not.[1]

Figure 2 maps state outcomes on the latest recreational cannabis ballot. Recreational cannabis remains a Schedule I controlled substance under the federal Controlled Substances Act. 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 commits new within-state indoor cultivation capacity to the local grid, and the vote sets the timing.

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, and all of them 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 of the country. In the restructured states customers may choose a competitive supplier, and many stay with the local distribution company (Borenstein and Bushnell, 2015). Across both structures, adding new generation or transmission takes years. Only 322 miles of transmission at 345 kilovolts and above were completed in

[1]State law sets the window for signature collection, which reaches two years in some states, and signatures are filed with the secretary of state months before the vote. Once certified, the measure is placed on a statewide ballot, in most states the next regular general election.

2024, against nearly 4,000 miles in 2013 (Shreve et al., 2025). Interconnection from request to commercial operation took nearly five years for projects built in 2023, against three years in 2015 (Rand et al., 2024). Where transmission constraints limit imports, short-run supply expansion cannot absorb a localized step increase in electricity demand, and it shows up in higher local prices.

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. 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. Where a regional market makes capacity payments, they 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 reaches local retail prices over several years, mainly through energy costs and the rate cases that pass them through.

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. A developer comparing counties learns how much local capacity is already committed from that study and from the load queues utilities publish, so a county whose capacity is being committed appears constrained years before the constraint reaches a retail rate. Grid capacity has limited new load in recent episodes. 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 generator retirements and to a jump in projected peak load in its footprint (PJM Interconnection, 2024). The independent market monitor puts the data center load included in that auction's forecast at 4,654 megawatts above existing levels, from new and expanding facilities (Monitoring Analytics, 2025). 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 continued into 2026.

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 gives 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. Aggregated to the county-year level over 2000-2024, the records give 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, taken from the Aterio classification and cross-checked by hand against publicly disclosed hyperscaler portfolios for the largest operators.

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 Amend-

ment 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. Dropping the Florida ballot leaves the decline in the data center stock at 0.70 facilities per county, within the range that Appendix Table C11 reports when each ballot is dropped in turn.

Maryland and New Jersey legalized through legislative referrals tied to a ballot question, not 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 in states that never held 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. 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. Surface thermal anomalies come from Landsat 8 thermal infrared band 10 at 100-meter native resolution resampled to 30 meters, with a 16-day revisit cycle. We use the band 10 surface temperature from the U.S. Geological Survey (USGS) Collection 2 Level-2 product. 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 the rooftop data begin in 2017.

We combine the three signals using inverse-variance weights estimated on state-licensed cultivation facilities in Colorado, Oregon, and Washington. The 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. An any-detection classifier marks a county-year with any triconfirmed pixel and splits counties in the mechanism analysis.

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. Where a sample split needs years before 2017, we instead split counties at the sample mean of a two-signal nightlight-thermal measure that drops the rooftop signal and extends back to 2000, using DMSP-OLS nightlight and MODIS land surface temperature before 2012, so both sides of each split are measured the same way. Appendix A documents the three-signal 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 and screens which

other-state counties enter a ballot's grid. The full electricity panel covers 2000-2024.

3.5 Labor Markets

Labor-market outcomes used in the alternative-mechanism tests come from 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. QCEW covers 95 percent of U.S. wage and salary employment. 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 and 541513 for data centers, NAICS 111419 and 111998 for cannabis cultivation, and NAICS 2382 for the building trades. 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 report them beside the county aggregate, which is complete in every county-year and provides the falsification test.

County Business Patterns (CBP) is 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. 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 center developers invest. This makes the legalization indicator endogenous to the very forces that drive siting, and a simple regression confounds the two. We address 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 shifts whether the state legalizes recreational cannabis cultivation. The discontinuity at the majority 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 simple-majority threshold

$$R_s = YesShare_s - 0.50 \tag{1}$$

and passage as $Pass_s = \mathbf{1}\{R_s \geq 0\}$. Throughout this section and the mechanism analysis of Section 5, 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. Counties in barely-passing states host 1.36 fewer active data centers 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 makes every estimate more precise.[2] Appendix Table C3 reports the pooled estimates. Counties in barely-passing states host 1.46 fewer active data centers, install 19 fewer megawatts of capacity, take in fewer new entries, and host fewer generation-equipped facilities.

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, not an effect of passage. Figure 3 shows a pre-ballot level difference between passing and failing states. The contrast is already open before the ballot, with passing states hosting 0.8-1.0 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. Because the cross-sectional design blends this pre-existing difference with the effect of passage, it overstates the causal effect. The pre-ballot gap moves little from year to year, and differencing removes a fixed difference. We therefore use 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

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

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, not its level. We keep observations within eight years of each ballot. Stacked designs bound the event horizon around each event in this way (Cengiz et al., 2019; Wing et al., 2024).[3]

[2] Because this cross-section is measured one year after each ballot 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.

[3] Each stacked ballot contributes counties from a single state, so $Pass_{sa} \times Post_{sta}$ is constant within a ballot-year and would be absorbed by year-by-ballot fixed effects. We therefore use calendar-year fixed effects alongside the county-by-ballot fixed effects. These absorb shocks common to every ballot in a year but keep the variation across ballots that identifies β, since in any calendar year some ballots have already passed, others have failed, and others have not yet been held.

Identification now rests on a parallel-trends assumption, conditional on the running-variable polynomial. The design allows a pre-existing discontinuity at the threshold, provided the discontinuity would not have changed had the ballot failed. Section 4.2 documented a gap 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 by dropping every control county served by a utility that also serves a passing state, which leaves the estimate essentially unchanged, as reported in 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. Appendix Table C1 reports the static specification, where the active data center stock falls by 0.702 facilities per county. 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, and new entries fall by 0.089 per year in the same direction. At the construction cost of 9.5-13.3 million dollars per megawatt reported in Section 2.1, the lost 16.2 MW is over 150 million dollars of investment that a county in a narrowly passing state does not receive. This forgone investment is not an unambiguous welfare 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 provides.

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. Caughey and Sekhon (2011) find that bare winners and losers of close U.S. House races differ on pre-determined characteristics, though such imbalance is the exception 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. Differencing does not address sorting across the threshold, which 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. Appendix Figure B1 plots the estimated density with its confidence bands.

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).[4] As Appendix Table C5 shows, only RGGI membership moves after passage, by 6 percentage points. The right-to-work and renewable-portfolio coefficients are within two hundredths of zero. Retail-choice status does not change in any bandwidth state during the window, so that coefficient is a mechanical zero, and the corporate rate moves by less than a percentage point. The estimates provide little support for a bundled-policy explanation.

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 could account for the post-ballot decline. The drift is small relative to the ballot-year change. The pre-ballot coefficients approach zero by 0.024 facilities a year, and the stock falls by 0.64 facilities in the ballot year alone, 27 times the annual drift. 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 Figure 5 shows it admits zero only for larger differences. The drift also reflects which ballots reach the farthest leads. The

[4]We select these as the state policies that could accompany a cannabis ballot while also shifting data center location through data centers' main input costs, namely 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.

panel begins in 2000 and ends in 2024, so each point of Figure 4 rests on a different set of ballots. Figure 6 holds that set fixed, on the twelve ballots observed in every year from five before to five after their own, and the pre-ballot coefficients are jointly indistinguishable from zero with the largest at 0.060 facilities. The effect on the same sample is 0.709 fewer facilities, against 0.702 in the full panel.

The control group is itself partly treated because three control states legalize within their own comparison windows. California legalizes in 2016 after its 2010 ballot failed, Oregon in 2014 after its 2012 ballot, and Arizona in 2020 after its 2016 ballot. 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 decline at 0.46 fewer facilities, and at 0.56 when South Dakota's overturned 2020 initiative is also counted as legalization. Appendix Table C6 reports both. At a three-year window the restricted and baseline samples give nearly identical estimates, 0.192 against 0.196, and they diverge only at the longest horizons, where the excluded states supply three of the six control ballots.

We next vary the event window, 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, as Appendix Table C7 shows. A wider window takes in more post-ballot years of a persistent effect, and it reaches the longer horizons that only the earlier ballots supply. The estimate is likewise stable across bandwidths, staying at 0.64-0.84 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. Appendix Table C8 reports the full bandwidth path.[5] It is also stable across 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, as Appendix Table C9 shows. The estimate is similar across ballots. Appendix Table C11 estimates a separate effect for each ballot-year cohort against the same pooled control group, and all eight are negative and significant, at 0.62-0.97 fewer facilities per county. Dropping one of the twelve ballot years at a time holds the pooled estimate at 0.56-0.96 fewer facilities, and dropping one ballot at a time at 0.53-0.84, significant in all twenty-six. 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, as Appendix Table C10 shows. We go no higher than a quadratic, since high-order polynomials are unreliable in this design (Gelman and Imbens,

[5]The optimal bandwidth of 0.04 leaves only four or five ballots on each side of the cutoff, too few for reliable inference, so we hold the baseline at 0.10, which keeps twelve failing and fourteen passing ballots.

2019).

A county's own supply of electricity could shape the prospects of both industries, so we let it shift the comparison. Utility-scale generating capacity sited in the county in the year before its own ballot, total and renewable, and the capacity it already had waiting in the interconnection queue, are each interacted with the post-ballot indicator and then with the full set of calendar years, which lets counties differing in supply follow different paths. The decline stays at 0.536-0.616 fewer facilities across the four specifications, against 0.702 without them. Appendix Table C12 reports them. Capacity is measured before the ballot because generation responds to the same demand shock, and interconnection requests are themselves an outcome in Section 5.3. Conditioning on either contemporaneously would net out part of the mechanism the paper is measuring. The design compares twelve failing and fourteen passing ballots, and with so few clusters the conventional cluster-robust standard errors can over-reject (Cameron and Miller, 2015). We therefore also report the wild cluster bootstrap, which stays reliable when clusters are few (Cameron et al., 2008), including the null-imposed variant (MacKinnon and Webb, 2018). Table C13 collects the estimate under analytical and bootstrap inference, clustered by ballot and by state, and the active-stock estimate remains distinguishable from zero under each procedure at the baseline bandwidth.

5 Mechanism

5.1 The Cannabis Cultivation First Stage

We first examine how legalization changes the electricity demand that cultivation places on the local grid. 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.[6] Detected cultivation appears in a small fraction of county-years, and a typical county-year is at zero, so the event study estimates the post-ballot shift in the distribution, not a change at the median. Figure 7 shows the first stage. Detected cultivation rises from the ballot year onward, with the per-year effect climbing from 0.096 square kilometers at passage to 0.176 square kilometers five years after the ballot. Against a sample mean of 0.0485 square kilometers, the post-ballot increase is 3.6 times the average detected area. Because our measure requires nightlight, thermal, and rooftop confirmation, the expansion

[6] A post-ballot rise in detected area could also reflect a detector that improves 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.

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 a concentrated local load.

5.2 The Electricity-Market Channel

We next examine whether cultivation raises local electricity prices. 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. We then 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 relates the price response to realized cultivation intensity.[7]

Within county, retail electricity prices rise with detected cultivation. In Table 1, the coefficient on the inverse hyperbolic sine of detected cultivation area is 0.051 for the log average retail price in Column (1), 0.046 for the log commercial price in Column (2), and 0.070 for the log industrial price 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. Table 2 shows a similar pattern in the ballot design. The cultivation-intensity interaction is positive in every column, at 2.2 percent per square kilometer of detected cannabis for the average price, 2.3 percent for commercial customers, and 2.6 percent for industrial customers. Legalized counties with more cultivation thus face larger post-ballot price increases, which ties the price increase to the counties where cultivation appears. A statewide tariff change would have moved prices in every county of a legalizing state, cultivating or not.

Panel B shows no detectable change in aggregate electricity consumption. Commercial consumption is the exception, rising 3.7 percent per square kilometer of detected cannabis. Commercial service is the margin on which licensed grow operations themselves buy power. Figure 8 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. Commercial buildings use 12.6 kilowatt-hours of electricity per square foot a year on average (U.S. Energy Information Administration, 2022), so at the eightfold intensity of Section 2.2 the fifth-year cultivation increase of 0.176 square kilometers, or 1.9 million

[7]Detected cultivation is itself a post-ballot outcome, so the interaction is not an exogenously assigned dose. All columns of Table 2 use the 2017-2024 detection sample, and Appendix Table C14 also reports the estimates on the full sample.

square feet, implies an average electricity demand of 22 MW, comparable to the 16.2 MW of data center capacity lost. The same calculation gives 12 MW in the ballot year, while the data center stock falls by 0.64 facilities in that year itself. The timing is consistent with a response to expected as well as realized grid conditions, which Section 5.3 examines.

Appendix Tables C14, C15, C16, and C17 report the corresponding estimates for the full panel, for cultivation area in levels, and for the eight-year window. The detection measure begins in 2017, so the within-county and cultivation-intensity estimates use 2017-2024 county-years, and the interaction is defined only on that sample. On the full panel the ballot-price estimate is nearly unchanged. The within-county estimates remain positive using cultivation area in levels and at the eight-year window, and at that window total consumption shows no detectable change in either the static or the dynamic design, as Appendix Figure B2 shows.

5.3 Anticipatory Responses

Data center siting is decided on horizons of a decade or more, so operators may respond to the resource constraint they expect legalization to create as well as to the price they observe today. A response to current prices alone would be concentrated where power has already become more expensive. A response to expected scarcity would also appear in counties whose own prices do not move. We split both the price effect and the data center effect on the scale of cannabis supply already present before the ballot. Pre-ballot cultivation proxies for places where post-legalization cultivation can expand more readily, so the energy cost that post-ballot cultivation drives should be concentrated where pre-ballot supply was large.

The decline in data centers is largest where cannabis supply was already established, and so is the price increase. Figure 9 splits on satellite-detected cultivation before the ballot, and Figure 10 on pre-ballot arrests for cannabis sale and manufacture.[8] The post-ballot price increase is larger in the high-supply group on both classifiers. The decline in data centers is larger where pre-ballot supply was higher, and the low-supply counties also lose data centers even without a measurable price increase. This pattern is more consistent with a forward-looking response than with a response to current retail prices alone. The pattern is similar using the year before the ballot, the average over available pre-ballot years, and a per-capita splitter.

Interconnection-queue applications provide an additional forward-looking outcome. The literature treats a filing as a signal of developer interest and attrition as a matter of transmis-

[8]The arrest measure reaches back to the 1970s and so classifies all 26 ballots, including the two 2012 legalizers that the three-signal measure cannot see before 2017.

sion capacity and connection cost. Johnston et al. (2023) separate the interconnection-cost channel from the wholesale-price channel and find little effect of a new generator beginning operation on wholesale prices at its location in PJM. Limited transmission capacity raises the network-upgrade costs facing new requests, which lifts withdrawal rates and may deter developers from filing at all (Rand et al., 2025). Queue applications fall with cultivation exposure on the local utility grid, including cultivation outside the county itself. Appendix Table C23 reports county-year counts of solar and gas generator interconnection-queue applications and dated withdrawals from the LBNL Queued Up records, in the panel difference-in-discontinuities of Section 4. Cultivation is measured only from 2017, so these estimates use the cultivation-dose subsample. The response scales with cultivation intensity. A square kilometer of cultivation anywhere on the county's utility grid lowers applications by 0.042. Splitting that dose into own-county and grid-neighbor components, applications fall by 0.199 per square kilometer in the county itself and by 0.029 per square kilometer elsewhere on the same grid. Withdrawals do not respond on either margin. Only a project that applied can withdraw, so the fall in applications also thins the pool at risk of withdrawing. A generator pays the network upgrades its own connection triggers, so capacity claimed anywhere on the shared network raises the cost of joining it.

5.4 Grid Transmission and Spillovers

We next examine the spatial extent of the price response by widening the unit from the county to the electricity grid. The grid is also where we can test the SUTVA assumption of 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, keeping an out-of-state county only when its balancing authority also operates in that 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 where the ballot does not legalize cultivation, whose composition is summarized in Table 3. The price response is concentrated within the utility's own service territory. In Table 4, an additional square kilometer of cultivation in a county's own territory raises its retail price by 0.83 percent, and the same square kilometer elsewhere on the same utility inside the state raises it by 0.78 percent. A single tariff covers the whole service territory. Across the state line the estimate falls to 0.37 percent. Table 6 reports the own, within-state and cross-state estimates with their confidence intervals and minimum detectable effects on the utility and balancing-authority grids. On the wider balancing-authority grid both the within-state and cross-state coefficients are much smaller, because the balancing authority averages over many

distant counties that share no local grid constraint, so we take the utility estimate as the primary measure. Each state's public utility commission decides rate cases for service inside its own borders, so the retail rate a customer pays is set on that customer's own side of the line, even where the two sides share a grid.

A county in another state enters a ballot's grid only if its balancing authority also operates in the balloting state. A shared utility alone does not qualify it, because a utility whose territories lie on separately operated systems would otherwise place counties on a grid they cannot reach. Tables C18 and C19 report both sets of results under a rule that asks only for a shared utility and under a rule that asks for a shared state border instead. The data center estimates and the within-state price estimate are similar under all three definitions. Form 861 reports revenue and sales by utility and state, so a county's price is a rescaling of the prices of the utilities serving it, and Table C21 reports how far that goes. Table C20 re-estimates the dose with the utility-state as the unit of analysis, using reported utility-state revenue and sales directly, without county-level allocation, and finds 0.66 percent per square kilometer against the 0.78 percent of the county specification. Residential rates move with it, at 0.62 percent, which a shift in the mix of customer classes could not produce, because cultivators buy commercial and industrial service.

Data center activity also declines in other-state counties on the same grid. Table 5 reports a fall of 0.36 facilities in legalizing counties that cultivate, 0.41 in legalizing counties that do not, and 0.45 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. The other-state group loses data centers on a similar scale, and the interconnection queue of Section 5.3 provides further evidence on the forward-looking channel.[9] Interference across the shared grid could move the estimate in either direction. A data center displaced from a legalizing state may relocate to a control county, while a decline that reaches control counties on the same grid would work the other way. Table C22 drops every control county served by a utility that also serves a state passing a ballot within the event window, and the data-center estimate moves from -0.702 to -0.735. All twelve failing ballots survive the removal at every horizon, so the movement comes from the dropped counties.

5.5 Labor-Market Channel

Labor competition is a natural alternative explanation. An expanding cultivation sector could raise local wages, and data centers facing a higher wage bill could forgo investment

[9]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 another state's ballot, not a standing level difference.

there. We examine it using industry scale, occupational overlap, wages, and employment. Both industries account for very small shares of county employment and draw on different occupations. Appendix Table C24 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 C25 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. On the cultivation side, employment in Support Activities for Crop Production 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 limited to local installation work, because the equipment itself is not made locally.

In Table 7 no wage column rises. The county-total weekly wage falls 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.[10] Wages in the broad sectors of the local economy show no increase either. Appendix Table C26 reports the weekly-wage effect in construction, manufacturing, retail, information, accommodation and food, leisure, and professional services, and no sector shows the positive response labor competition requires. The employment effects in the county total, data centers, cannabis cultivation, and those contractors are indistinguishable from zero. Figure 11 traces total county employment in the dynamic design and shows a flat path with no break at the ballot. The data center wage declines after the ballot. Figure 12 shows flat pre-ballot coefficients followed by a post-ballot decline of 220 dollars a week that later attenuates. Labor competition would push this wage up, as data centers pay more to keep their workers, whereas a decline is consistent with fewer data center entries. Among county-years with an operating cultivator, the indoor cultivation wage rises by 619 dollars a week after legalization, in Appendix Table C27, reflecting higher labor demand in the sector that legalization expands. We find little evidence that labor

[10]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.

competition explains the decline in data center activity.

5.6 Other Channels

We also consider land, capital, and knowledge spillovers. Data centers site on industrial and commercial parcels chosen for proximity to high-voltage substations and fiber backbone, where land is a small share of a capital outlay dominated by the shell, power, and cooling. Licensed indoor cultivation occupies warehouse space under the security and building-code requirements that Sections 2.2 and 2.3 describe. 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 cannabis cultivation in a county does not compete for the funds that build them. Hyperscaler capital expenditure reaches hundreds of billions of dollars a year and is allocated across the country by a handful of firms. Cannabis cultivation and data centers share little common technology or knowledge base through which one could raise the other's productivity, and the two do not compete in the same product markets (Bloom et al., 2013; Moretti, 2021). Indoor growing is a horticultural and climate-control activity, and data centers are computing infrastructure. Wages and employment also show no broad increase across local sectors, from construction and manufacturing to retail, information, and professional services.

6 Location Choice of Data Centers

6.1 Model and Estimation

We use a discrete-choice model of data center siting to translate the reduced-form electricity-price response into spatial reallocation and welfare magnitudes. For scale, a 100 MW 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, so a price increase of several percent adds millions of dollars a year over a facility's multi-decade life. The model treats each prospective data center as a single discrete choice over U.S. counties and recovers a price coefficient that we use to compute counterfactual spatial reallocation and welfare costs. 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.[11] Conditional on the choice, the facility proceeds through the construction, activation, and

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

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. A price wedge in one county lowers its siting probability, and under this restriction the probability it loses spreads over the remaining counties in proportion to their existing shares. The restriction fixes where mobile capacity goes when one county becomes more expensive, and it is why Section 6.2 interprets the counterfactual as a direction and not as a facility-by-facility relocation.

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. The residual $\widehat{v}_c$ enters the utility in equation (6) and absorbs the correlation between price and unobserved attractiveness. The identifying assumption is that close-margin ballot passage is as good as randomly assigned and shifts data center siting only through the local electricity price. Fiber backbone, spare substation capacity, and serviced industrial land all draw data centers and all raise the local price. They enter the price residual, and under this assumption the control function addresses the resulting price

endogeneity.[12]

Table 8 reports the estimates. The price coefficient is -5.05, 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. The price coefficient is identified from the siting choices through the cannabis control function, so the model translates the ballot variation into a price elasticity and a spatial counterfactual. A 12.1 percent legalization-induced price increase lowers a county's relative siting probability by 44 percent.

6.2 Counterfactual

We construct a counterfactual that removes the legalization-induced price increase in counties of states where legalization took effect. We measure capacity as operating capacity at active facilities, which totals 36 gigawatts nationally in 2024 and aligns with published U.S. data center power demand (Shehabi et al., 2024). We build the counterfactual price in each such 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 local entry the reduced form also documents.

Relative to the counterfactual, legalization moves capacity away from the legalizing states toward states that did not legalize by ballot. Appendix Figure B3 maps the direction of the state-level shift, with Texas, Pennsylvania, and Illinois the largest gainers and the legalizing states with the largest existing base the largest losers.[13] The gainers follow from the model's proportional reallocation, which directs removed capacity toward the largest existing bases outside the legalizing states, as Section 6.1 sets out. We interpret the counterfactual as a direction and not as 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. Appendix Table C28 looks outside the model at state trade in the products data centers are built from, classified by Waugh (2026). Ballot-passing states import fewer graphics processors, servers, and storage units relative to their other trade, and the gap holds for servers alone, the most data-center-specific import.

[12] Section 4.4 finds no jump at the cutoff in right-to-work status, renewable-portfolio standards, retail choice, or the corporate tax rate.

[13] The three gaining states did not legalize through a ballot, so the price wedge the counterfactual removes does not apply to them.

6.3 Welfare

The higher retail price is borne by local electricity consumers in legalizing states. Industrial and commercial users tied to local production, residential households, and public facilities pay the higher post-legalization retail rate. 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 legalizing states that the design identifies, a cost borne in those states, not in states that legalized well away from the threshold. The calculation counts only the retail-price wedge and excludes the property-tax base, local employment, and agglomeration economies that the lost capacity would have supported. Applied to captive 2024 consumption in the counties whose states fall inside the estimation bandwidth, the wedge costs 12.2 billion dollars a year across 581 counties, or 21.0 million dollars per county. Appendix Figure B4 gives the loss by state, concentrated in the largest legalizing electricity markets. A ballot-passing county also forgoes more than 150 million dollars of facility investment in shell, power, and cooling, before any computing hardware.

7 Conclusion

Close recreational cannabis ballots add indoor cultivation to local electricity grids, and data center activity declines where they pass. After a ballot passes, the active data center stock falls by 0.702 facilities per county and operating capacity by 16.2 MW. Five years after the ballot, the added cultivation implies an average electricity demand of 22 MW per county, comparable to the data center capacity lost. In legalizing counties, retail prices rise 2.2 percent more for each square kilometer of detected cultivation and 2.6 percent more for industrial users, while total electricity consumption does not move measurably. We find little evidence for labor, land, or capital channels.

The decline in data centers begins in the ballot year, before cultivation has scaled, and it also appears in counties in other states that share the grid. Interconnection-queue applications fall where cultivation expands. These patterns are consistent with siting that responds to expected grid scarcity over the long horizons on which data centers are planned. The location-choice model implies that without legalization more capacity would be built in the narrowly passing states, and that the higher retail prices cost consumers there over 12

billion dollars a year. The same grid lock-in may matter for electricity-intensive industries, including cryptocurrency mining, semiconductor fabrication, and battery and electrolyzer production.

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.

Alvarez, Fernando E., David Argente, Joyce Chow, and Diana Van Patten, "Data Centers and Local Economies in the Age of AI: A Shift-Share Approach," Working Paper 35194, National Bureau of Economic Research May 2026.

Ancona, Camilla, Ofek Lauber Bonomo, Anton Rozhkov, and Maurizio Porfiri, "Urban Infrastructure and Fossil Fuel Industrial Legacy Drive US Data Center Siting," *Nature Cities*, 2026, *3*, 756–767.

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.

Benetton, Matteo, Giovanni Compiani, and Adair Morse, "When Cryptomining Comes to Town: High Electricity-Use Spillovers to the Local Economy," Working Paper 31312, National Bureau of Economic Research 2023.

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.

Colas, Mark and Emmett Reynier, "Optimal Subsidies for Residential Solar," *Journal of Political Economy Microeconomics*, 2026, *4* (1), 213–264.

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

Denver Department of Public Health and Environment, "Cannabis Environmental Best Management Practices Guide: Energy," Technical Report, City and County of Denver, Denver, Colorado October 2019.

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.

Du, Rui, Xiaojin Sun, and Siqi Zheng, "Bits and Mortar: The Microspatial Economics of Data Centers," Research Paper 26/05, MIT Center for Real Estate 2026.

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 2026: Updated Scenarios of U.S. Data Center Electricity Use and Power Strategies," White Paper 3002034696, Electric Power Research Institute, Palo Alto, CA February 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.

Johnston, Sarah, Yifei Liu, and Chenyu Yang, "An Empirical Analysis of the Interconnection Queue," Working Paper 31946, National Bureau of Economic Research December 2023. Revised August 2026.

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.

MacKinnon, James G. and Matthew D. Webb, "The Wild Bootstrap for Few (Treated) Clusters," *The Econometrics Journal*, 2018, *21* (2), 114–135.

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.

_____, "Energy-Intensive Indoor Cultivation Drives the Cannabis Industry's Expanding Carbon Footprint," *One Earth*, 2025, *8* (3), 101179.

Monitoring Analytics, "Analysis of the 2025/2026 RPM Base Residual Auction, Part G," Technical Report, Monitoring Analytics, Independent Market Monitor for PJM June 2025.

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, Steven Zhang, Chris Talley, Will Gorman, Ryan Wiser, Joachim Seel, Julie Mulvaney Kemp, Seongeun Jeong, and Fritz Kahrl, "Queued Up: 2025 Edition, Characteristics of Power Plants Seeking Transmission Interconnection As of the End of 2024," Technical Report, Lawrence Berkeley National Laboratory, Berkeley, California December 2025.

———, ———, **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, Lawrence Berkeley National Laboratory, Berkeley, California April 2024.

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 LBNL-2001637, Lawrence Berkeley National Laboratory, Berkeley, California December 2024.

Shreve, Nathan, Zach Zimmerman, Gretchen Kershaw, and Rob Gramlich, "Fewer New Miles: Strategic Industries Held Back by Slow Pace of Transmission," Technical Report, Grid Strategies July 2025.

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.

Suárez Serrato, Juan Carlos 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.

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* (7), 644–650.

Turner & Townsend, "Data centre construction cost index 2025-2026," Technical Report, Turner & Townsend, Leeds, United Kingdom November 2025.

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

U.S. Energy Information Administration, "2018 Commercial Buildings Energy Consumption Survey, Table C14: Electricity Consumption and Expenditure Intensities, 2018," Technical Report, U.S. Energy Information Administration, Washington, DC December 2022.

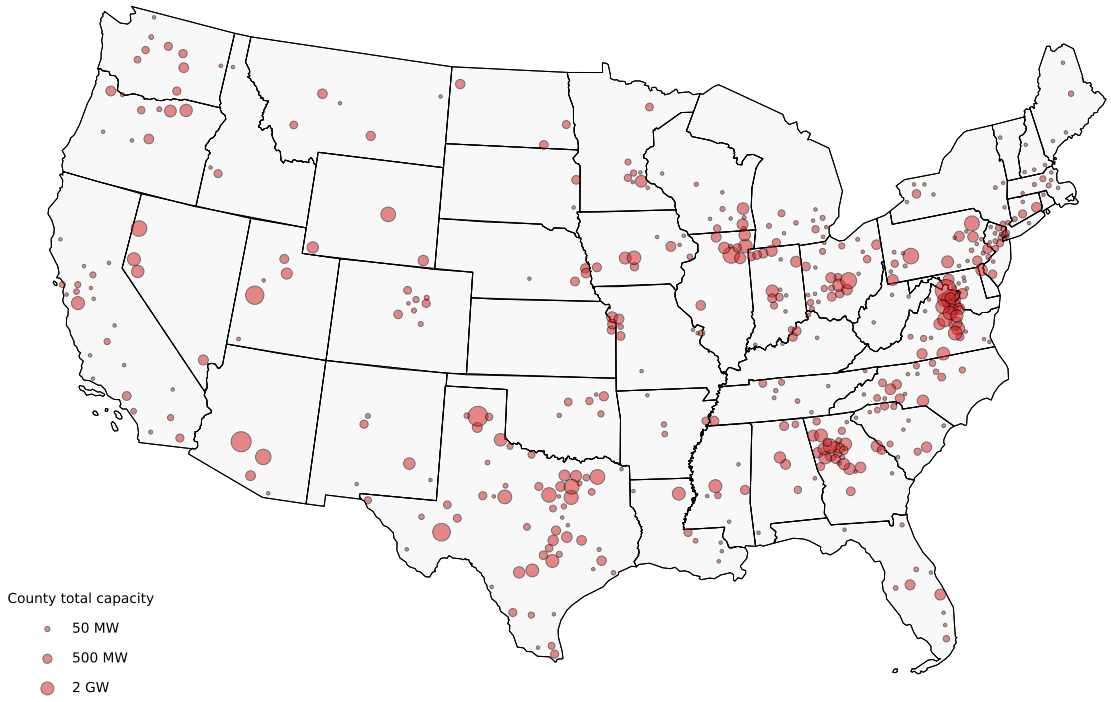
Waugh, Michael E., "Trade in AI-Related Products," Working Paper 35053, National Bureau of Economic Research April 2026.

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.

Wing, Coady, Seth M. Freedman, and Alex Hollingsworth, "Stacked Difference-in-Differences," Working Paper 32054, National Bureau of Economic Research 2024.

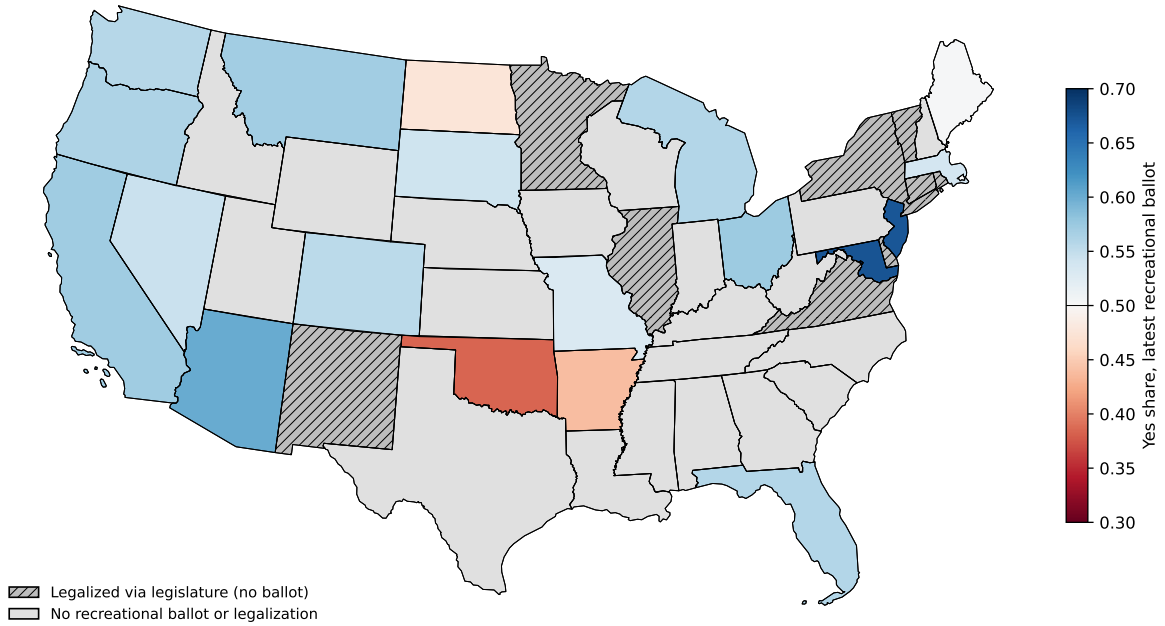
Zhang, Calvin, "AI Data Centers and Electricity Prices: Much Ado About Something?," Working Paper 6930258, SSRN 2026.

Figure 1. U.S. data center locations



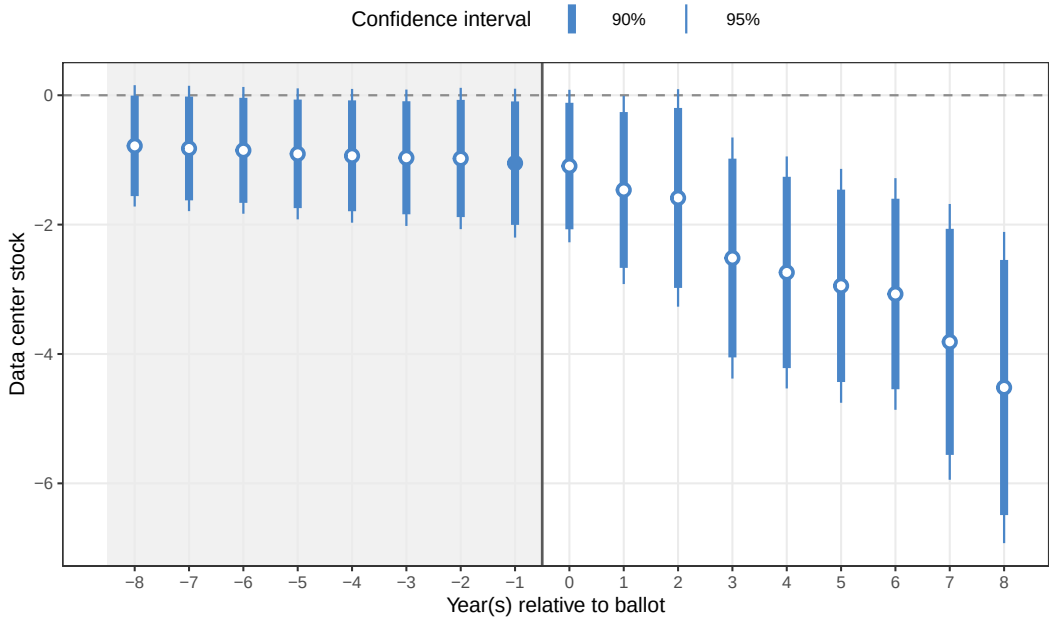
Notes: Each red marker is a county centroid sized by the total nameplate power capacity of data center facilities located in that county. The sample includes 5,140 facilities across 461 counties and pools all development stages, from announcement to operation. Marker area is proportional to the square 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



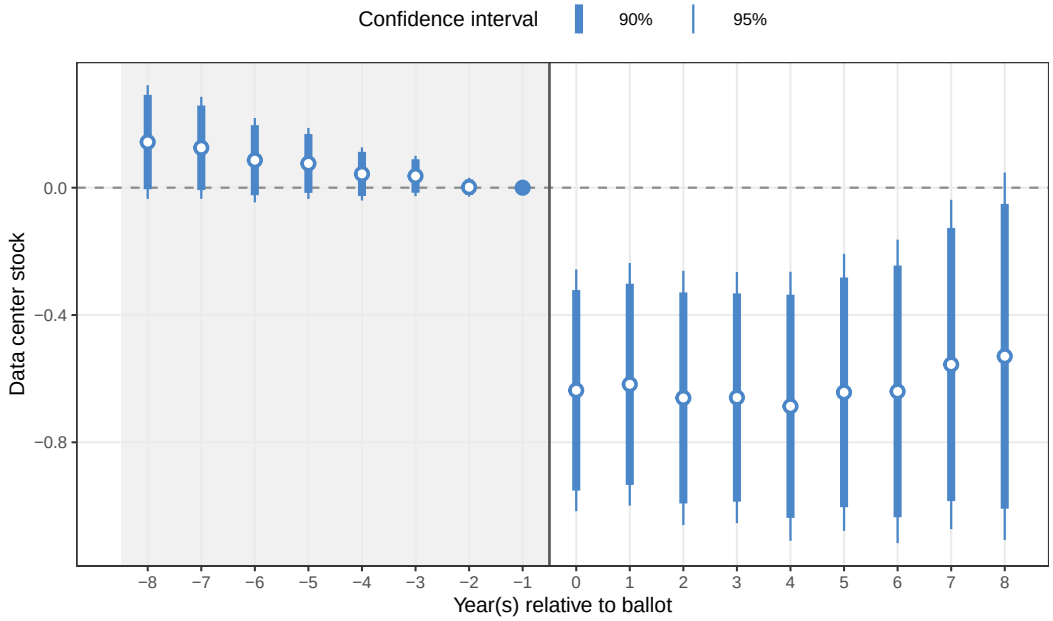
Notes: States are shaded by the Yes share on the most recent recreational cannabis ballot, with passing ballots taking precedence over later defeats. The diverging color scale is centered at the 50 percent simple-majority threshold, so blue shades indicate a Yes majority and red shades a No majority. States that legalized recreational use through legislative action rather than ballot are shaded gray with diagonal hatching. States with no recreational ballot and no recreational legalization are shaded light 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



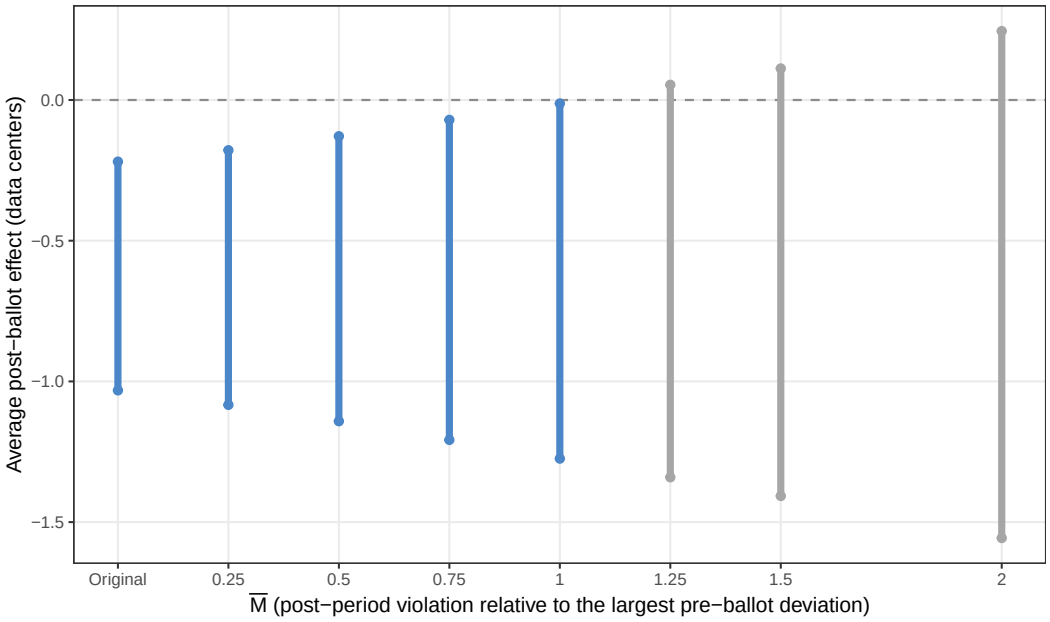
Notes: The figure plots the pooled cross-sectional RD coefficient on *Pass*, estimated separately at each event time k. Bars show 90 and 95 percent confidence intervals, based on standard errors clustered at the ballot level. The estimation sample is restricted to $|R_{sa}| \leq 0.10$. The outcome has a mean of 0.446 active facilities per county-year.

Figure 4. Data center stock around cannabis legalization



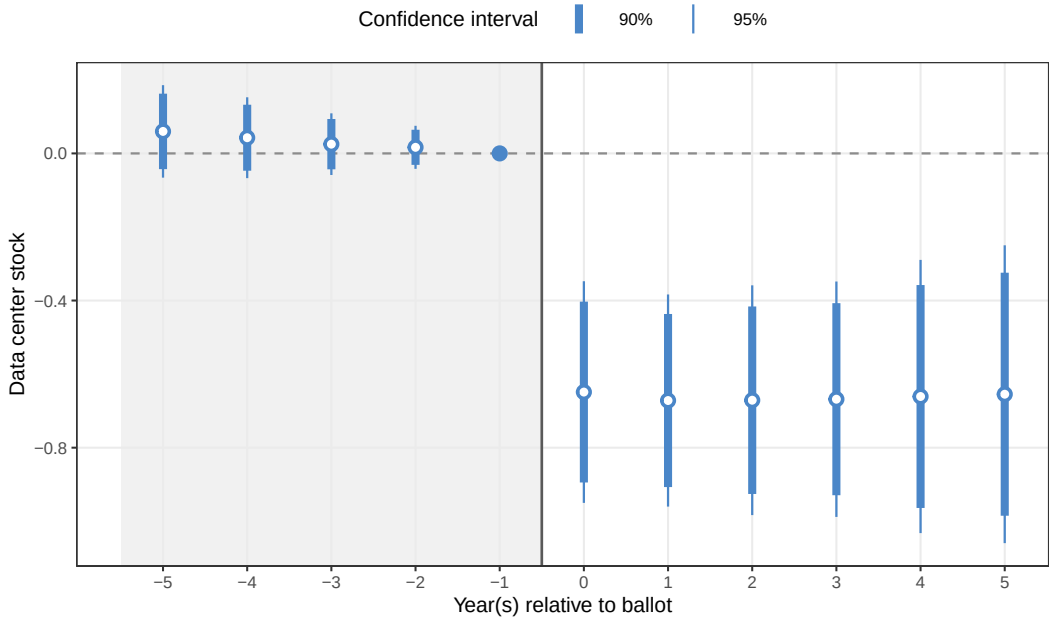
Notes: The figure reports a panel difference-in-discontinuities event study in which the single post-ballot term is replaced by interactions of *Pass* with event time, and the year before the ballot is the omitted reference period. All specifications include county-by-ballot and calendar-year fixed effects. Bars show 90 and 95 percent confidence intervals, 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



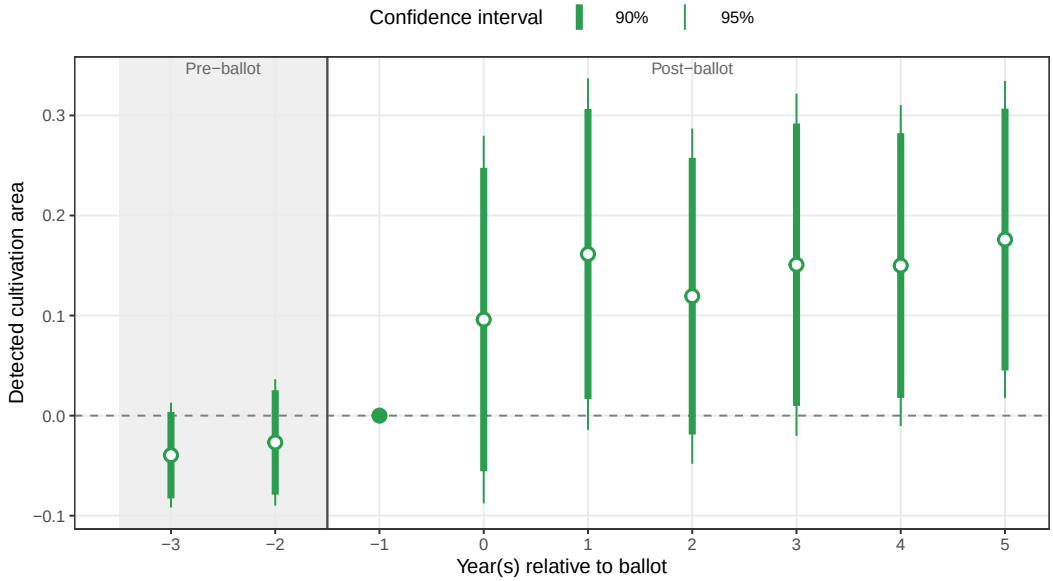
Notes: The figure reports the Rambachan and Roth (2023) relative-magnitudes robust 95 percent confidence set for the average post-ballot effect on the active data center stock, as the permitted post-period violation of parallel trends grows from the original confidence interval, the $\bar{M} = 0$ case in which the restriction imposes exact parallel trends after the ballot, to $\bar{M} = 2$, twice the largest violation observed across the seven pre-ballot years. The average is taken over the nine post-ballot event-study coefficients. Intervals that exclude zero are shown in blue and those that admit zero in gray. The event study is the panel difference-in-discontinuities specification of Figure 4, with standard errors clustered at the ballot level and the sample restricted to $|R_{sa}| \leq 0.10$.

Figure 6. Data center stock around cannabis legalization, balanced five-year window



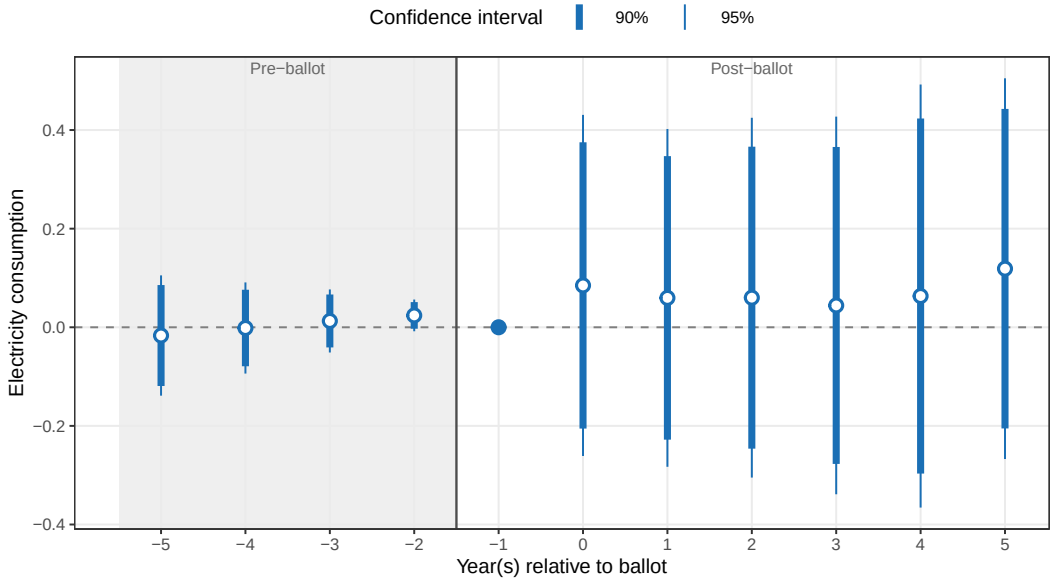
Notes: The sample is the 12 ballots whose counties are observed in every year from five before to five after their own ballot, so the set of ballots behind each point is identical and movement across event time cannot be composition. Nine of these ballots passed and three failed, and the sample covers 409 counties and 5,324 county-years. Arizona's 2016 ballot and Oregon's 2012 ballot are excluded from the control group because Arizona legalized in 2020 and Oregon in 2014, inside their own windows. The legalization dates are those of Section 4.4, taken from every passing ballot at any margin. Each point is the passing-minus-failing gap in the active data center stock at that year relative to the ballot, with the year before the ballot as the reference. Bars show 90 and 95 percent confidence intervals from the clustered-t distribution, based on standard errors clustered at the ballot level. The specification includes county-by-ballot and calendar-year fixed effects, and the estimation sample is restricted to $|R_{sa}| \leq 0.10$.

Figure 7. Detected cannabis cultivation around legalization



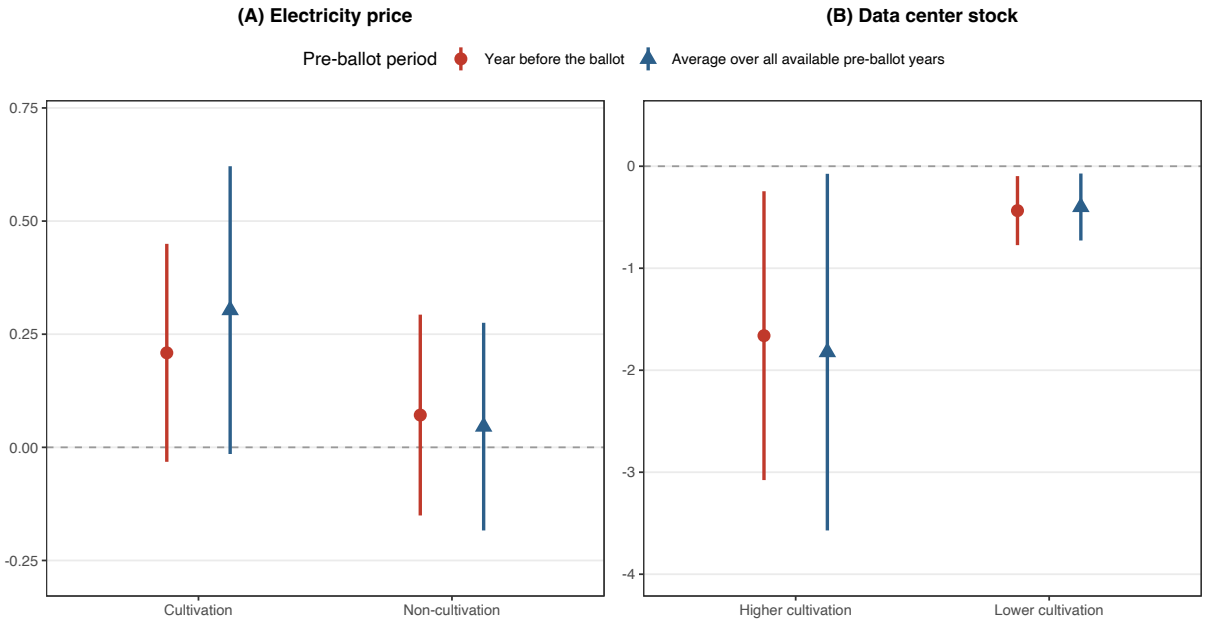
Notes: The figure reports a panel difference-in-discontinuities event study for cultivation area detected by all three signals in square kilometers, estimated over a plus or minus five-year window, and the year before the ballot is the omitted reference period. Bars show 90 and 95 percent confidence intervals, based on standard errors clustered at the ballot level. The estimation sample is restricted to $|R_{sa}| \leq 0.10$ and contains 6,927 county-year observations. The display is trimmed to three years before the ballot; the regression uses the full plus or minus five-year panel. The outcome has a mean of 0.0485 square kilometers per county-year.

Figure 8. Total electricity consumption around legalization



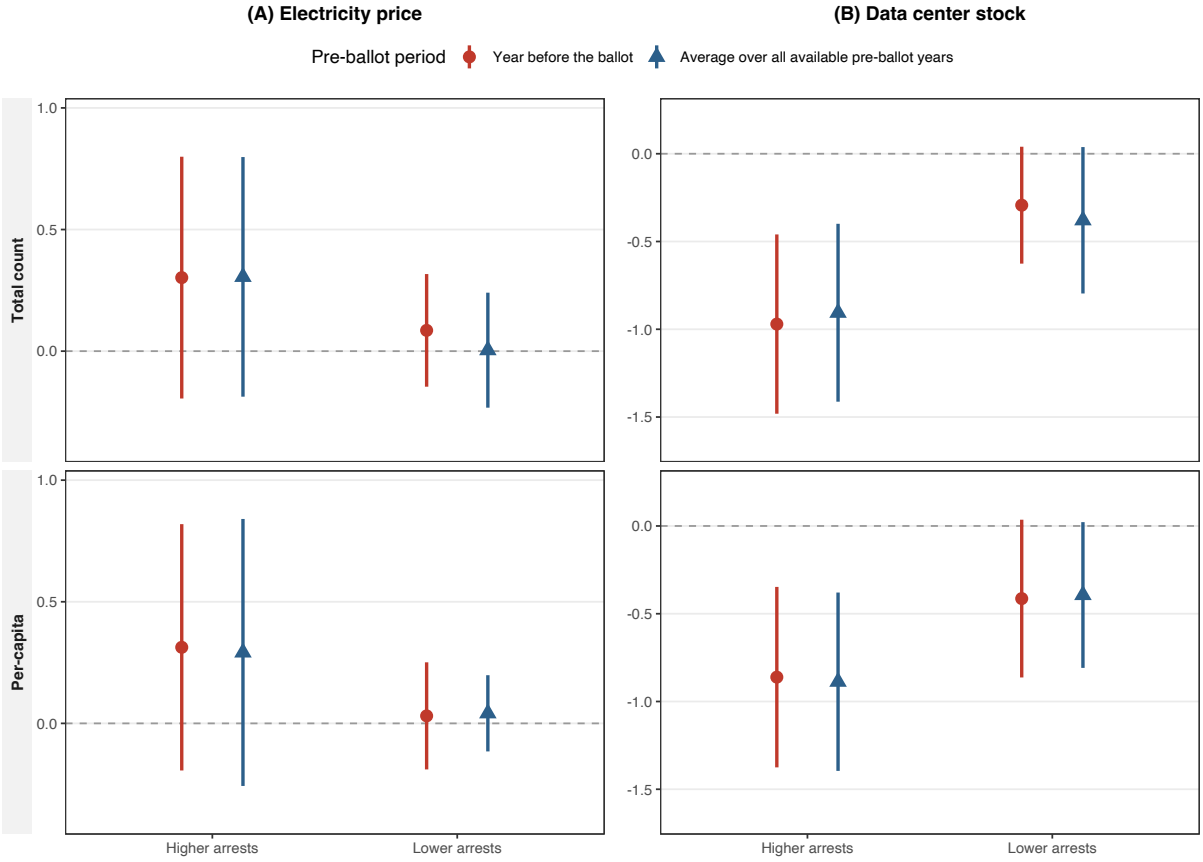
Notes: The figure reports a panel difference-in-discontinuities event study for population-weighted total electricity consumption in inverse hyperbolic sine of megawatt-hours, estimated over a plus or minus five-year window, and the year before the ballot is the omitted reference period. Bars show 90 and 95 percent confidence intervals, based on standard errors clustered at the ballot level. The estimation sample is restricted to $|R_{sa}| \leq 0.10$ and contains 11,563 county-year observations. 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 9. Price and data center responses by pre-ballot cultivation exposure



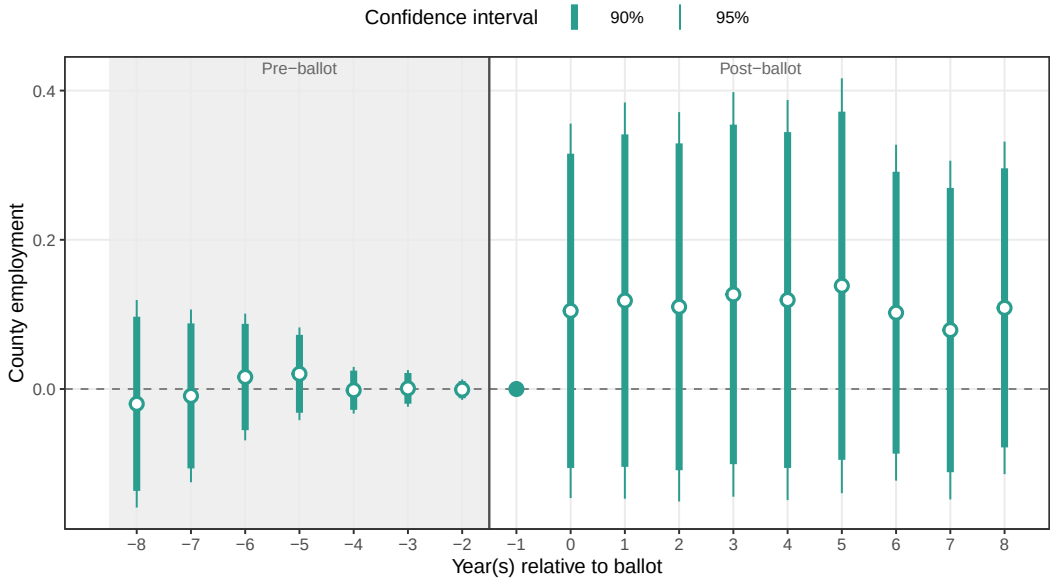
Notes: The figure reports the difference-in-discontinuities Pass $\times$ Post effect on the average retail electricity price (Panel A, log) and the active data center stock (Panel B, levels), with counties split by pre-ballot satellite-detected cultivation status. Panel A uses the three-indicator any-detection 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, and each column has its own vertical axis. The two series within each panel use the cultivation classifier in the year before the ballot $(t-1)$ and the average over all available pre-ballot years, shown as red circles and blue triangles respectively. 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 95 percent confidence intervals use standard errors clustered at the ballot level.

Figure 10. Price and data center responses by pre-ballot cannabis-supply enforcement



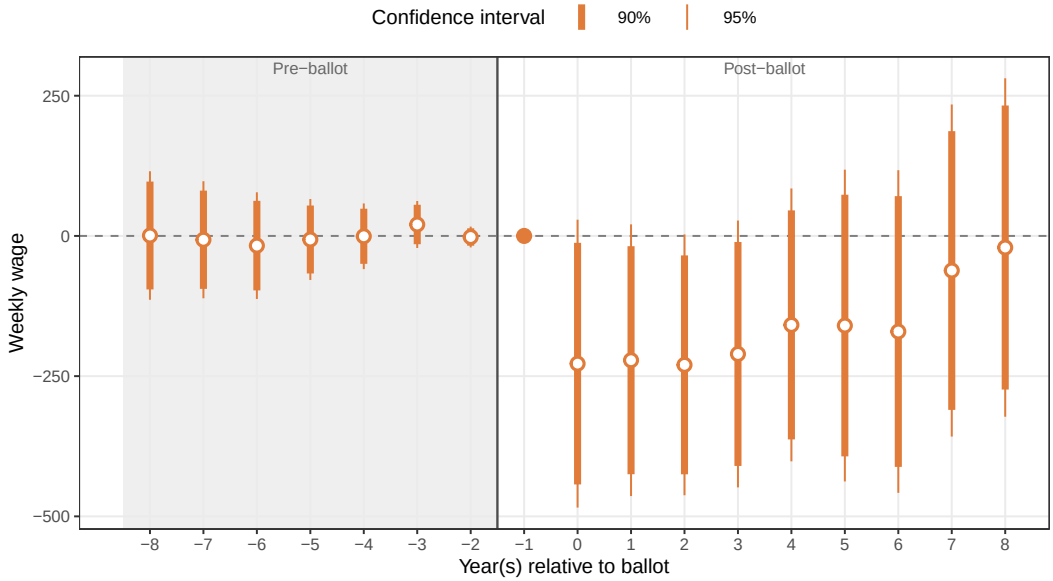
Notes: The figure reports the difference-in-discontinuities Pass × Post effect on the average retail electricity price (Panel A, log) and the active data center stock (Panel B, levels), with counties split by pre-ballot cannabis-supply enforcement, namely marijuana sale and manufacturing arrests from the UCR Arrests by Age, Sex, and Race series (Kaplan, openICPSR #102263). Unlike the satellite cultivation measure in Figure 9, 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, and the bottom row uses the per-capita rate. Counties are split at the sample-wide median within each splitter, each column has its own vertical axis, shared between the two splitter rows, and the two series within each cell use the splitter in the year before the ballot $(t-1)$ and the average over all available pre-ballot years, shown as red circles and blue triangles respectively. 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 95 percent confidence intervals use standard errors clustered at the ballot level.

Figure 11. 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. Bars show 90 and 95 percent confidence intervals, based on standard errors clustered at the ballot level. The estimation sample is restricted to $|R_{sa}| \leq 0.10$ and contains 16,870 county-year observations. 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. Missing county-years are set to zero before the inverse-hyperbolic-sine transform. See Table 7.

Figure 12. Data center sector wages around legalization



Notes: The figure reports a panel difference-in-discontinuities event study for data-center-sector wages in dollars per week, 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. Bars show 90 and 95 percent confidence intervals, based on standard errors clustered at the ballot level. The estimation sample is restricted to $|R_{sa}| \leq 0.10$ and contains 16,870 county-year observations. 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 of detected cultivation area (km^2)	0.0508*** (0.0071)	0.0457*** (0.0134)	0.0700*** (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 relationship 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 detected cultivation area in km^2 from the three-signal satellite measure. 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-signal 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 and balancing-authority grid footprints

Diagnostic	Utility grid	Balancing-authority 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)		
In-state, cultivating (county-ballots)	112	112
In-state, non-cultivating (county-ballots)	609	609
Other state, same grid (county-ballots)	386	4,177

Notes: The table reports the composition of the spillover sample on the utility and balancing-authority grids. Each grid footprint is constructed for the 26 ballots within the bandwidth $|R_{sa}| \leq 0.10$ as the union of counties served by a grid unit, a utility or a balancing authority, that also serves the balloting state. The treated-group counts are post-ballot county-ballots, split into in-state cultivating counties, in-state non-cultivating counties, and counties in other states on the same grid. The two in-state groups are the same on both grids by construction. Rest-of-grid cultivation is summed over the other counties in the grid-unit footprint.

Table 4. Spillover of cannabis cultivation onto the county electricity price

Grid	Utility		Balancing authority	
Spillover	Pooled	Split	Pooled	Split
	(1)	(2)	(3)	(4)
Variables				
Pass×Post	0.0364	0.0454	-0.0137	-0.0141
	(0.0908)	(0.0890)	(0.0231)	(0.0241)
Own-county km^2×Pass×Post	0.0096***	0.0083**	0.0040	0.0040
	(0.0034)	(0.0037)	(0.0029)	(0.0030)
Pooled spillover km^2×Pass×Post	0.0062***		0.0005**	
	(0.0020)		(0.0002)	
Within-state km^2×Pass×Post		0.0078**		0.0004
		(0.0030)		(0.0003)
Cross-state km^2×Pass×Post		0.0037*		0.0005**
		(0.0020)		(0.0002)
Fixed effects				
County × ballot	Yes	Yes	Yes	Yes
Calendar year	Yes	Yes	Yes	Yes
Fit statistics				
Observations	14,630	14,630	88,561	88,561

Notes: The table reports panel difference-in-discontinuities estimates of the spillover of cannabis cultivation onto the county electricity price, at the county-ballot-year level. The outcome is the log population-weighted average retail price. Own-county cultivation and rest-of-grid cultivation, both in km^2, each enter interacted with Pass×Post. The pooled columns enter rest-of-grid cultivation as one term, and the split columns separate it into a within-state and a cross-state component. An out-of-state county counts as a grid neighbor only if its balancing authority also operates in the balloting state. In-state counties are unaffected by this rule. The rule is redundant on the balancing-authority footprint, where the condition holds by construction. 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 sample is restricted to 2017-2024, the period over which the three-signal satellite measure is available. 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 5. Spillover of cannabis cultivation onto the county data-center stock

Grid	Utility		Balancing authority	
Spillover	Pooled	Split	Pooled	Split
	(1)	(2)	(3)	(4)
Variables				
In-state, cultivating	-0.3671*	-0.3584*	-0.1361	-0.1318
	(0.1848)	(0.1796)	(0.0973)	(0.0929)
Pooled spillover	-0.4291**		-0.1999**	
	(0.1610)		(0.0772)	
In-state, non-cultivating		-0.4085**		-0.1724
		(0.1569)		(0.1012)
Other state, same grid		-0.4497**		-0.2009**
		(0.1863)		(0.0766)
Fixed effects				
County × ballot	Yes	Yes	Yes	Yes
Calendar year	Yes	Yes	Yes	Yes
Fit statistics				
Observations	25,308	25,308	134,358	134,358

Notes: The table reports panel difference-in-discontinuities estimates of the spillover of cannabis cultivation onto the county data-center stock, at the county-ballot-year level. The outcome is the active data-center count in levels. The three county groups are in-state cultivating counties, in-state non-cultivating counties, and counties in other states on the same grid. The pooled spillover combines the in-state non-cultivating and other-state groups. An out-of-state county counts as a grid neighbor only if its balancing authority also operates in the balloting state. In-state counties are unaffected by this rule. The rule is redundant on the balancing-authority footprint, where the condition holds by construction. 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. 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 6. Own, within-state, and cross-state spillovers onto the retail electricity price

Grid	Dose	Estimate (%)	SE (%)	95% CI (%)	p-value	MDE (%)
Utility	Own county	0.830**	0.373	[0.098, 1.562]	0.037	1.046
Utility	Within-state	0.778**	0.295	[0.200, 1.356]	0.015	0.826
Utility	Cross-state	0.374*	0.200	[-0.019, 0.766]	0.076	0.561
BA	Own county	0.403	0.301	[-0.187, 0.992]	0.194	0.842
BA	Within-state	0.041	0.028	[-0.013, 0.096]	0.148	0.077
BA	Cross-state	0.049**	0.019	[0.012, 0.085]	0.016	0.052

Notes: The table reports the continuous dose specification of Table 4 on the utility and balancing-authority (BA) footprints, at the county-ballot-year level. Own cultivation, cultivation elsewhere on the grid inside the state, and cultivation on the grid across the state line each enter interacted with Pass×Post, alongside the running-variable polynomial. The outcome is in logs, so the coefficient times one hundred is the percent change per square kilometer of cultivation. 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 percent power. 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 sample is restricted to 2017-2024, the period over which the three-signal satellite measure is available. Standard errors are clustered at the ballot level. Three, two, and one asterisks denote statistical significance at the 1, 5, and 10 percent levels, respectively.

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 at the county-year level, with running-variable controls and the fixed effects listed in the table. 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 16,870-observation panel difference-in-discontinuities sample enters every column. Columns (1) and (5) are the NAICS 10 county total, where wage is the employment-weighted county-average and employment is total headcount. Columns (2) and (6) are broad data centers in NAICS 518210 and 541513, Columns (3) and (7) cannabis cultivation in NAICS 111419 and 111998, and Columns (4) and (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, dollars per week for wage and workers for employment. 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.

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 2015-2024, each choosing from a sampled choice set of fifty counties, the chosen county plus 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 surface temperature from Landsat 8 thermal infrared band 10, 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 surface-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.

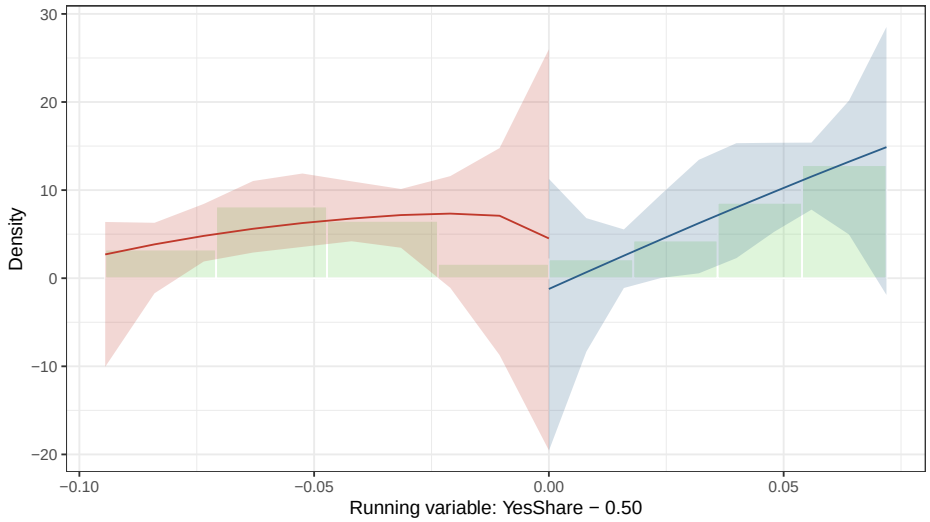
The third signal is optical rooftop detection from Sentinel-2 surface reflectance at 10-meter resolution, distributed through the Copernicus program with a five-day revisit and available from 2017. Greenhouse and warehouse roofs are bright and spectrally flat, so we combine a brightness index across 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 on 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 area under the receiver operating characteristic curve of 0.87, compared with 0.76 for nightlight alone, 0.78 for thermal anomalies alone, and 0.82 for rooftop classification alone. The analysis uses the county-year measures defined in Section 3.3, namely the area of pixels on which all three signals exceed their detection thresholds, the composite activity index, an indicator for any such pixel in the county-year, and an indicator for two-signal detected area above the sample mean. The geographic distribution of the resulting high-confidence measure is mapped in Appendix Figure B5.

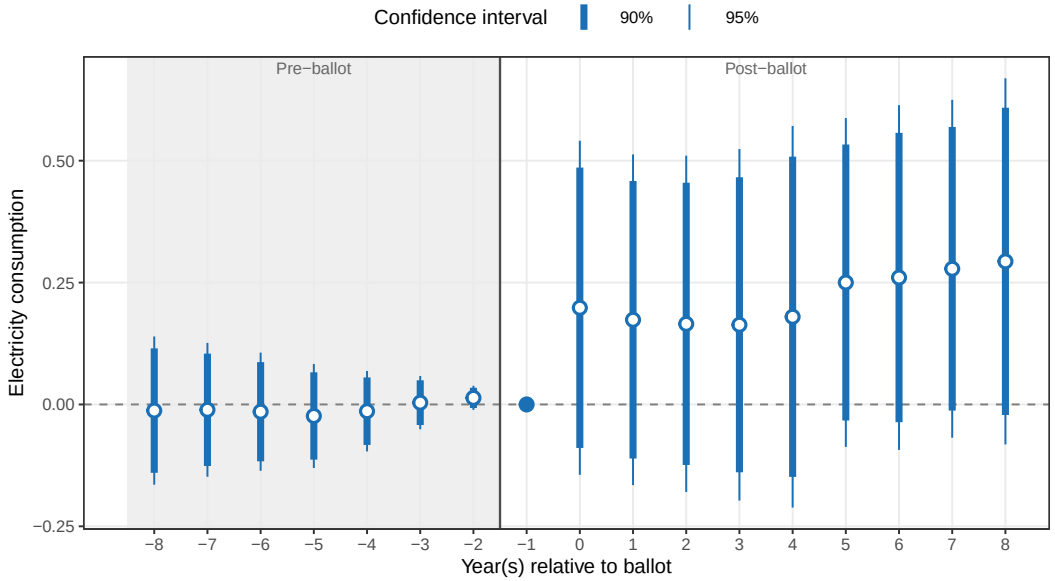
B Additional Figures

Figure B1. Density of the running variable at the cutoff



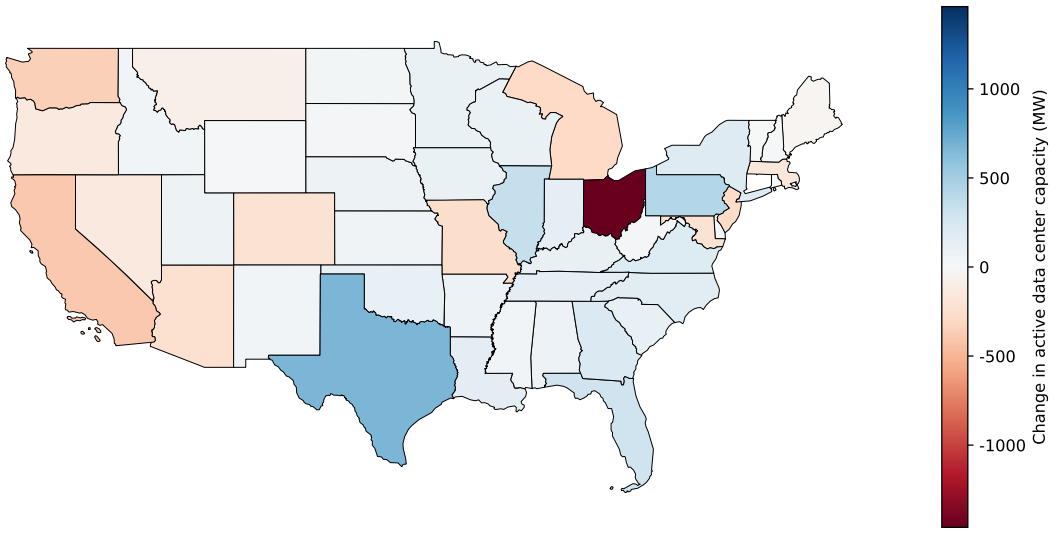
Notes: The figure plots the local-polynomial estimate of the density of the running variable R_{sa} on each side of the fifty-percent cutoff, with 95 percent confidence bands, following Cattaneo et al. (2020a). The estimation bandwidth is $|R_{sa}| \leq 0.10$.

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



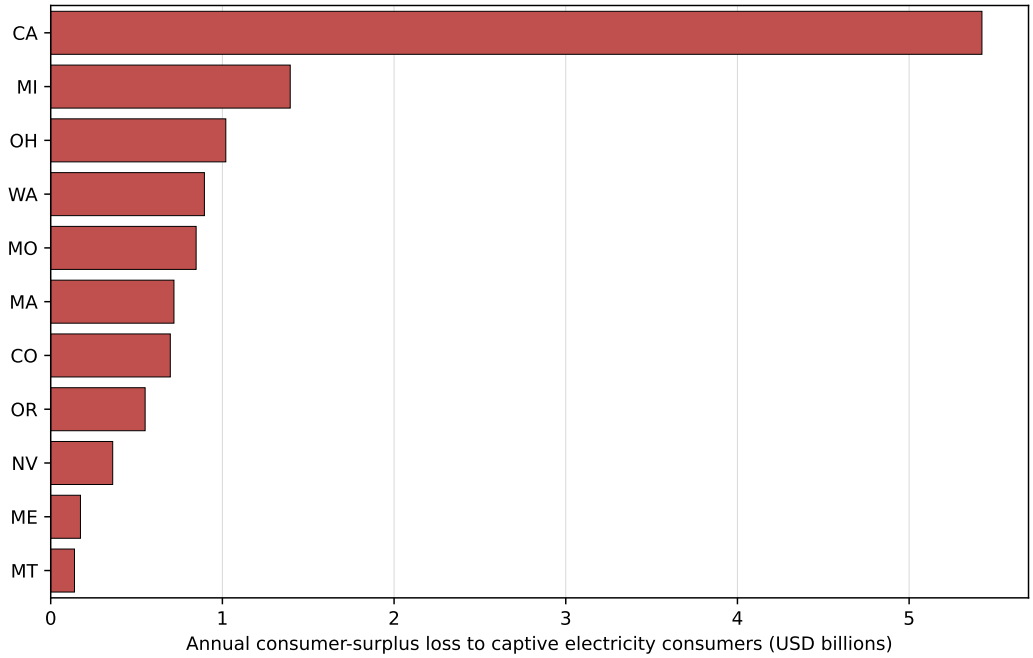
Notes: The figure repeats Figure 8 at a plus or minus eight-year window. The year before the ballot is the omitted reference period, bars show 90 and 95 percent confidence intervals based on standard errors clustered at the ballot level, and the estimation sample is restricted to $|R_{sa}| \leq 0.10$ and contains 16,633 county-year observations.

Figure B3. Reallocation of U.S. data center capacity from recreational cannabis legalization



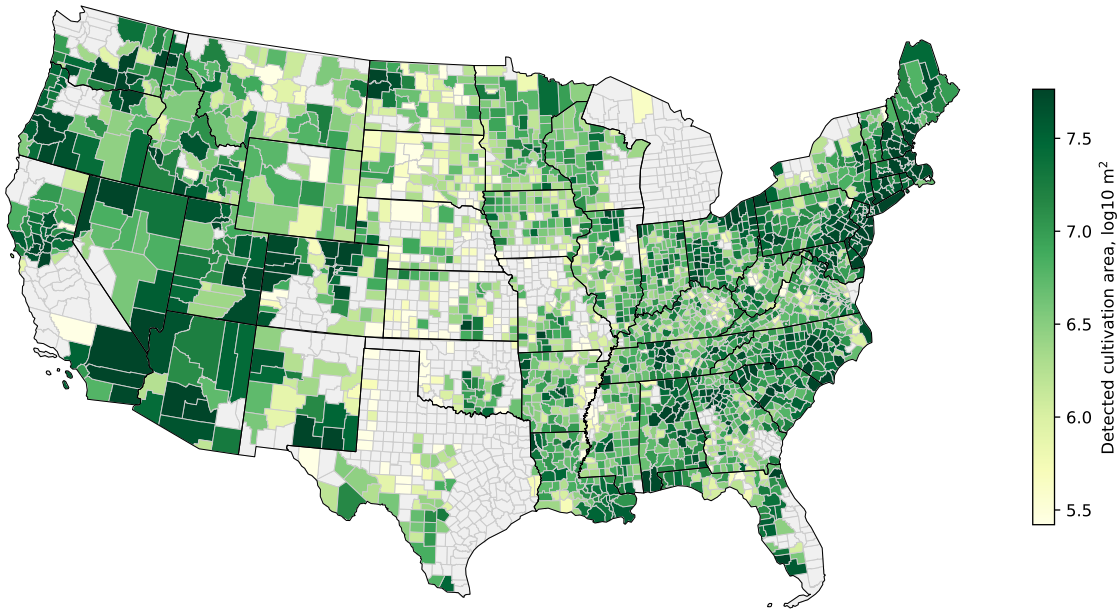
Notes: States are shaded by the change in 2024 active data center capacity that the model attributes to legalization, measured as modeled capacity under observed prices minus modeled capacity under the counterfactual prices, holding the observed national total fixed. Blue shades mark states whose capacity is higher because of legalization elsewhere, the destinations of the reallocation, and red shades mark legalizing states whose capacity the price wedge lowers. 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 legalizing 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 12.2 billion dollar bandwidth total reported in Section 6.3. 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. Counties with zero detected cultivation are shown in light gray. 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. The data center stock effect across the three designs

Outcome:		Stock	
Single-ballot cross-section: Pass	-1.355		
	(0.8572)		
Pooled cross-section: Pass		-1.464**	
		(0.7014)	
Panel difference-in-discontinuities: Pass×Post			-0.7019***
			(0.2194)
Fixed effects			
County × ballot			Yes
Calendar 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

Notes: The table reports the effect of recreational cannabis legalization on the active data center stock under three designs, using the 2000-2024 county-year data center panel. The single-ballot cross-section keeps each state's most recent ballot and has one observation per county. The pooled cross-section stacks every ballot event and has one observation per county and ballot. Both cross-sections measure the outcome one year after the ballot and include no fixed effects. The panel difference-in-discontinuities is estimated at the county-ballot-year level over a window of plus or minus eight years around the ballot and includes county-by-ballot and calendar-year fixed effects. The estimation bandwidth is $|R_{sa}| \leq 0.10$. 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 C2. Single-ballot cross-sectional estimates by data center outcome

Outcome:	Stock	Entry	Stock (MW)	Entry (MW)	Without on-site generation	With on-site generation	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

Notes: The table reports single-ballot cross-sectional regression discontinuity estimates of the effect of recreational cannabis legalization on data center outcomes, with one observation per county. Each state enters with its most recent ballot, and outcomes are measured one year after the ballot, using the 2000-2024 county-year data center panel. The columns report the active stock, new entries, operating and entrant capacity in megawatts, the active stock without and with on-site generation, and the pipeline stock. The specification includes no fixed effects. The estimation bandwidth is $|R_{sa}| \leq 0.10$. Standard errors are clustered at the ballot level, with 13 clusters, and are reported in parentheses. Three, two, and one asterisks denote statistical significance at the 1, 5, and 10 percent levels, respectively.

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

Outcome:	Stock	Entry	Stock (MW)	Entry (MW)	Without on-site generation	With on-site generation	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

Notes: The table reports pooled cross-sectional regression discontinuity estimates of the effect of recreational cannabis legalization on data center outcomes. The pooled cross-section stacks every ballot event and has one observation per county and ballot. Outcomes are measured one year after the ballot, using the 2000-2024 county-year data center panel. The columns report the active stock, new entries, operating and entrant capacity in megawatts, the active stock without and with on-site generation, and the pipeline stock. The specification includes no fixed effects. The estimation bandwidth is $|R_{sa}| \leq 0.10$. Standard errors are clustered at the ballot level, with 23 clusters, and are reported in parentheses. Three, two, and one asterisks denote statistical significance at the 1, 5, and 10 percent levels, respectively.

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

Outcome:	Stock	Entry	Stock (MW)	Entry (MW)	Without on-site generation	With on-site generation	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 × ballot	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Calendar 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

Notes: The table reports panel difference-in-discontinuities estimates of the effect of recreational cannabis legalization on data center outcomes, at the county-ballot-year level. The sample is the 2000-2024 county-year data center panel, stacked by ballot and restricted to a window of plus or minus eight years around the ballot. The columns report the active stock, new entries, operating and entrant capacity in megawatts, the active stock without and with on-site generation, and the pipeline stock. All specifications include county-by-ballot and calendar-year fixed effects and control for $R_{sa} \times \text{Post}$ and $R_{sa} \times \text{Pass} \times \text{Post}$. The estimation bandwidth is $|R_{sa}| \leq 0.10$. Standard errors are clustered at the ballot level, with 26 clusters, and are reported in parentheses. Three, two, and one asterisks denote statistical significance at the 1, 5, and 10 percent levels, respectively.

Table C5. Covariate balance in the state regulatory environment

Outcome:	RPS (binding)	RGGI member	Right-to-work	Retail choice	Corporate tax
Pass×Post	-0.0160	0.0618*	-0.0157	0.0000	-0.8835
	(0.0910)	(0.0351)	(0.0576)	(0.0000)	(0.8341)
Fixed effects					
County × ballot	Yes	Yes	Yes	Yes	Yes
Calendar year	Yes	Yes	Yes	Yes	Yes
Fit statistics					
Observations	16,870	16,870	16,870	16,870	16,870

Notes: Each column reports the Pass×Post coefficient from the baseline panel difference-in-discontinuities specification applied to the state-year regulatory covariate listed in the header, merged by ballot state onto the stacked county-ballot-year panel over 2000-2024. A cannabis ballot may signal voter willingness to adopt other regulation that also affects data center siting. If so, these covariates would jump at the cutoff after passage. RPS indicates that the state has a binding mandatory Renewable Portfolio Standard in that year, from DSIRE and NCSL. RGGI indicates that the state is a current participating member of the Regional Greenhouse Gas Initiative. Right-to-work indicates that the state has a right-to-work statute, from NCSL. Retail choice indicates that the state has full retail electricity choice, from EIA. Corporate tax is the top statutory state corporate income tax rate in percent, from the ITEP Historical State Tax Rate Data compiled from FTA. 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. 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 C6. Not-yet-treated control exclusion

Outcome	Active stock		Active entry	
Specification	Baseline	Excluding legalized	Baseline	Excluding legalized
$\hat{\beta}$: Pass×Post	-0.7019***	-0.4582**	-0.0886*	-0.0357
	(0.2194)	(0.1654)	(0.0447)	(0.0242)
Fixed effects				
County × ballot	Yes	Yes	Yes	Yes
Calendar 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

Notes: The table reports panel difference-in-discontinuities estimates of the effect of recreational cannabis legalization on the active data center stock and on new entries, at the county-ballot-year level. The sample is the 2000-2024 county-year data center panel, stacked by ballot and restricted to a window of plus or minus eight years around the ballot. The excluding-legalized sample drops each control state's observations from the year the state itself legalizes onward, which removes California from 2016, Oregon from 2014, and Arizona from 2020. South Dakota's overturned 2020 initiative and Florida's 2024 measure, which fell short of the required 60 percent supermajority, remain passing ballots and are not treated as legalization. Treating South Dakota's 2020 majority as legalization instead gives -0.556 for the stock and -0.046 for entries. At a three-year window the baseline and excluding-legalized stock estimates are -0.196 and -0.192. All specifications include county-by-ballot and calendar-year fixed effects and control for $R_{sa} \times \text{Post}$ and $R_{sa} \times \text{Pass} \times \text{Post}$. The estimation bandwidth is $|R_{sa}| \leq 0.10$. Standard errors are clustered at the ballot level, with 26 clusters, and are reported in parentheses. Three, two, and one asterisks denote statistical significance at the 1, 5, and 10 percent levels, respectively.

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

Notes: The table reports panel difference-in-discontinuities estimates of the Pass×Post coefficient on the active data center stock at three event windows, at the county-ballot-year level. The sample is the 2000-2024 county-year data center panel, stacked by ballot and restricted to the event window around each ballot. All specifications include county-by-ballot and calendar-year fixed effects. The estimation bandwidth is $|R_{sa}| \leq 0.10$. 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 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***
			(0.344)	(0.068)
0.05	6	6	-1.035**	-0.177*
			(0.393)	(0.085)
0.06	9	10	-0.787**	-0.117
			(0.312)	(0.068)
0.07	10	12	-0.741**	-0.113*
			(0.293)	(0.061)
0.08	11	14	-0.839***	-0.142**
			(0.285)	(0.061)
0.10	12	14	-0.702***	-0.089*
			(0.219)	(0.045)
0.12	13	15	-0.722***	-0.114**
			(0.232)	(0.052)
0.15	14	15	-0.642***	-0.088*
			(0.206)	(0.049)

Notes: The table reports panel difference-in-discontinuities estimates of the Pass×Post coefficient on the active data center stock and on new entries across estimation bandwidths, at the county-ballot-year level. The sample is the 2000-2024 county-year data center panel, stacked by ballot. 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. 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 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

Notes: 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, at the county-ballot-year level. The sample is the 2000-2024 county-year data center panel, stacked by ballot. The event window is plus or minus eight years around the ballot, and the estimation bandwidth is $|R_{sa}| \leq 0.10$. 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 C10. 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 × ballot	Yes	Yes	Yes	Yes
Calendar 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

Notes: The table reports panel difference-in-discontinuities estimates of the effect of recreational cannabis legalization on the active data center stock and on new entries, at the county-ballot-year level. The sample is the 2000-2024 county-year data center panel, stacked by ballot and restricted to a window of plus or minus eight years around the ballot. All specifications include county-by-ballot and calendar-year fixed effects and control for $R_{sa} \times \text{Post}$ and $R_{sa} \times \text{Pass} \times \text{Post}$. The quadratic specification adds $R_{sa}^2 \times \text{Post}$ and $R_{sa}^2 \times \text{Pass} \times \text{Post}$ to the running-variable polynomial. The levels of R_{sa} and R_{sa}^2 are constant within a ballot and are absorbed by the county-by-ballot fixed effects. The estimation bandwidth is $|R_{sa}| \leq 0.10$. Standard errors are clustered at the ballot level, with 26 clusters, and are reported in parentheses. Three, two, and one asterisks denote statistical significance at the 1, 5, and 10 percent levels, respectively.

Table C11. The data center effect by ballot-year cohort

Ballot Calendar year	Passing ballots	Treated counties	Data center stock
2012	2	103	-0.9174***
			(0.2159)
2014	2	65	-0.6592**
			(0.2750)
2016	4	105	-0.6157**
			(0.2388)
2018	1	83	-0.9744***
			(0.2503)
2020	2	122	-0.9325***
			(0.2543)
2022	1	115	-0.6804***
			(0.2162)
2023	1	88	-0.6877**
			(0.2558)
2024	1	67	-0.8867***
			(0.2395)

Notes: Each row reports the effect of passage for the ballots held in that year, estimated jointly in one regression in which Pass×Post is interacted with cohort indicators, so every cohort is measured against the same pooled control group of failing-ballot counties. A separate regression per cohort is not possible, because a within-year comparison needs both a passing and a failing ballot in that year, and 2014, 2020, and 2023 have no failing ballot while the 2002-2010 ballots include no passing one. Dropping one cohort at a time leaves the pooled estimate between -0.958 and -0.563, significant at five percent in all 12 runs. Dropping one ballot at a time leaves it between -0.839 and -0.530, significant in all 26. The sample is the 2000-2024 county-year data center panel, stacked by ballot at the county-ballot-year level and restricted to a window of plus or minus eight years around the ballot. The regression includes 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 and are reported in parentheses. Three, two, and one asterisks denote statistical significance at the 1, 5, and 10 percent levels, respectively.

Table C12. The data center effect conditioning on pre-ballot generation capacity

	Estimate	95% CI	N
No supply control (baseline)	-0.7019***	[-1.154, -0.250]	16,870
	(0.2194)		
+ existing and renewable capacity × Post	-0.6158***	[-0.994, -0.237]	16,870
	(0.1838)		
+ existing and renewable capacity × year	-0.5922***	[-0.975, -0.209]	16,870
	(0.1860)		
+ queued capacity as well, × Post	-0.5681***	[-0.927, -0.209]	16,870
	(0.1742)		
+ queued capacity as well, × year	-0.5365***	[-0.888, -0.185]	16,870
	(0.1708)		

Notes: The table reports panel difference-in-discontinuities estimates of the effect of recreational cannabis legalization on the active data center stock, at the county-ballot-year level, without and with controls for pre-ballot generation capacity. The sample is the 2000-2024 county-year data center panel, stacked by ballot and restricted to a window of plus or minus eight years around the ballot. Existing and renewable generation capacity are the utility-scale nameplate megawatts sited in the county in the year before its own ballot, from EIA Form 860M. Queued capacity is the megawatts of interconnection requests filed in the county before the ballot, from the Lawrence Berkeley National Laboratory queue data. All three are predetermined and enter as the inverse hyperbolic sine of megawatts, since 209 of the 891 bandwidth counties host no utility-scale generator at all. Conditioning on capacity as measured *at* the outcome year would be conditioning on a response to the same shock, since interconnection requests are themselves an outcome in Section 5.3. The interactions instead let the predetermined level shift the post-ballot break or the whole calendar-year path. Form 860 covers generators of one megawatt and above, so this is utility-scale capacity. All specifications include county-by-ballot and calendar-year fixed effects. The estimation bandwidth is $|R_{sa}| \leq 0.10$. 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 C13. Inference for the headline data center stock effect

Inference procedure	G	SE	p	95% CI
Analytical, cluster by ballot	26	0.219	0.004	[−1.154, −0.250]
Analytical, cluster by state	17	0.236	0.009	[−1.201, −0.202]
Wild bootstrap (unrestricted), ballot	26	0.219	0.028	[−1.317, −0.086]
Wild bootstrap (null-imposed), ballot	26	0.219	0.010	[−1.226, −0.177]
Wild bootstrap (null-imposed), state	17	0.236	0.023	[−1.307, −0.096]

Notes: The table reports inference for the panel difference-in-discontinuities effect of recreational cannabis legalization on the active data center stock, at the county-ballot-year level. The sample is the 2000-2024 county-year data center panel, stacked by ballot and restricted to a window of plus or minus eight years around the ballot. The specification includes county-by-ballot and calendar-year fixed effects, and the estimation bandwidth is $|R_{sa}| \leq 0.10$. The point estimate is $\widehat{\beta} = -0.702$ active data centers in every row, and only the inference procedure varies. Ballot clustering, with 26 clusters, is the primary specification. State clustering, with 17 clusters, pools the states appearing on more than one ballot and is a more conservative alternative. Analytical intervals are the cluster-robust t interval with $G-1$ degrees of freedom, where G is the number of clusters. Bootstrap intervals are symmetric studentized percentile-t intervals from $B = 4{,}999$ Rademacher draws (Cameron et al., 2008). "Unrestricted" draws residuals from the fitted model. "Null-imposed" draws them from the model with the treatment effect set to zero, the variant recommended by MacKinnon and Webb (2018).

Table C14. 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: The table reports panel difference-in-discontinuities estimates of the effect of recreational cannabis legalization on log electricity prices, 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 full sample uses all ballot-state counties in the bandwidth. The restricted sample covers 2017-2024, the period of satellite detection coverage, and retains 19 of the 26 ballots. Four ballots 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. 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 C15. Electricity prices and detected cultivation in levels

	Average (1)	Commercial (2)	Industrial (3)
$\hat{\gamma}$: Detected cultivation 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 relationship 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 detected cultivation area from the three-signal satellite measure. 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. Electricity prices and detected cultivation within county at the eight-year window

	Average (1)	Commercial (2)	Industrial (3)
$\hat{\gamma}$: Inverse hyperbolic sine of detected cultivation area (km^2)	0.0510*** (0.0092)	0.0477*** (0.0132)	0.0670*** (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 relationship 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 detected cultivation area in km^2 from the three-signal satellite measure. All specifications include county-by-ballot and calendar-year fixed effects. The estimation bandwidth is $|R_{sa}| \leq 0.10$, the event window is plus or minus eight years around the ballot, 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 C17. 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: 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. The specification follows Table 2 except for the event window. 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. Both panels use the sample restricted to 2017-2024, the period of satellite detection coverage, so the interaction columns are defined throughout. Standard errors are clustered at the ballot level 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 C18. Electricity-price spillover under the two alternative grid-footprint rules

Footprint rule	Any shared utility		Bordering state	
Spillover	Pooled	Split	Pooled	Split
	(1)	(2)	(3)	(4)
Variables				
Pass×Post	0.0485	0.0678	0.0527	0.0653
	(0.0912)	(0.0853)	(0.0916)	(0.0880)
Own-county km^2×Pass×Post	0.0072*	0.0037	0.0102**	0.0082*
	(0.0039)	(0.0040)	(0.0036)	(0.0041)
Pooled spillover km^2×Pass×Post	0.0046*		0.0065**	
	(0.0024)		(0.0023)	
Within-state km^2×Pass×Post		0.0090***		0.0089**
		(0.0031)		(0.0032)
Cross-state km^2×Pass×Post		0.0014		0.0030
		(0.0023)		(0.0024)
Fixed effects				
County × ballot	Yes	Yes	Yes	Yes
Calendar year	Yes	Yes	Yes	Yes
Fit statistics				
Observations	15,926	15,926	14,828	14,828

Notes: The table reports panel difference-in-discontinuities estimates of the spillover of cannabis cultivation onto the county electricity price under two alternative grid-footprint rules, at the county-ballot-year level. Columns (1)-(2) admit any out-of-state county sharing a utility with the balloting state, an alternative shared-utility rule that counts PacifiCorp's Utah counties as grid neighbors of California. Columns (3)-(4) instead require the county's state to border the balloting state, with borders derived from Census TIGER county boundaries. Table 4 reports the balancing-authority rule used in the text, and the outcome and dose terms are defined as there. 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. 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 C19. Data-center spillover under the two alternative grid-footprint rules

Footprint rule	Any shared utility		Bordering state	
Spillover	Pooled	Split	Pooled	Split
	(1)	(2)	(3)	(4)
Variables				
In-state, cultivating	-0.2934*	-0.2988*	-0.3377*	-0.3278*
	(0.1584)	(0.1661)	(0.1833)	(0.1796)
Pooled spillover	-0.3714**		-0.4093**	
	(0.1403)		(0.1625)	
In-state, non-cultivating		-0.3468**		-0.3760**
		(0.1387)		(0.1576)
Other state, same grid		-0.4194**		-0.4496**
		(0.1741)		(0.1953)
Fixed effects				
County × ballot	Yes	Yes	Yes	Yes
Calendar year	Yes	Yes	Yes	Yes
Fit statistics				
Observations	29,518	29,518	25,898	25,898

Notes: The table reports panel difference-in-discontinuities estimates of the spillover of cannabis cultivation onto the county data-center stock under two alternative grid-footprint rules, at the county-ballot-year level. The footprint rules are as in Table C18. Table 5 reports the balancing-authority rule used in the text, and the outcome and county groups are defined as there. The pooled spillover combines the in-state non-cultivating and other-state groups. 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. 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 C20. Cultivation spillover onto the utility-state electricity price

Footprint	Balancing authority			Any shared utility
Price	Average	Residential	Commercial	Average
	(1)	(2)	(3)	(4)
Variables				
Pass×Post	0.1067	0.1329	0.0629	0.0962
	(0.1810)	(0.1946)	(0.1712)	(0.1705)
Own utility-state km^2	0.0066***	0.0062***	0.0035**	0.0049*
	(0.0018)	(0.0018)	(0.0013)	(0.0024)
Same state, other utility km^2	-0.0003	-0.0004	-0.0004	0.0002
	(0.0017)	(0.0020)	(0.0018)	(0.0017)
Same utility, other state km^2	0.0022	0.0003	0.0020	0.0018
	(0.0030)	(0.0036)	(0.0034)	(0.0022)
Fixed effects				
Utility-state × ballot	Yes	Yes	Yes	Yes
Calendar year	Yes	Yes	Yes	Yes
Fit statistics				
Observations	3,421	3,421	3,421	3,724

Notes: The table reports panel difference-in-discontinuities estimates of the cultivation spillover onto the electricity price, with the utility-state as the unit of analysis. The outcome is the log price reported on EIA Form 861 for the utility-state-year, total revenue divided by total sales. No county-level allocation enters at any point, so the estimate is free of the construction described in Table C21. Doses are cultivation in square kilometers. The own dose covers the counties this utility-state serves. The same-state, other-utility dose covers a different utility in the same state, which shares neither revenue nor sales with this one. The same-utility, other-state dose covers the same utility across the state line. Counties are assigned to one utility-state by modal utility, so the three doses do not overlap. Columns (1)-(3) use the balancing-authority footprint rule of the main text, and Column (4) repeats the average-price specification under the alternative shared-utility rule. All specifications include utility-state-by-ballot and calendar-year fixed effects. The estimation bandwidth is $|R_{sa}| \leq 0.10$. 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. Allocation of utility prices to counties

Counties in the electricity panel	3,135
served by exactly one utility	256 (8.2%)
of those, share matching the reported utility price to within 0.1%	95.6%
Multi-county utility-state-years	7,117
share in which every county carries an identical price	5.8%
median within-cell coefficient of variation of price	0.039
R^2 of county log price on utility-state × year	0.935
R^2 of county log price on calendar year alone	0.342
In the spillover estimation sample	
county-ballot-year observations	14,630
distinct counties	1,125
distinct utility-states behind them	160
distinct ballots (clusters)	23

Notes: County electricity price is built by allocating each utility-state's revenue and its sales to counties with the same population weight and then dividing one by the other, so the weight cancels and a county served by a single utility inherits that utility's reported price exactly. The upper rows cover the counties of the electricity panel, and the lower rows cover the spillover estimation sample. Table C20 re-estimates the spillover at the utility-state level, with the allocation removed entirely.

Table C22. Headline estimates with grid-connected control counties removed

Horizon (years)	Controls dropped	Share of controls	Data-center stock	Log price
None	0	0%	-0.7019***	0.1832
			(0.2194)	(0.2076)
0	18	3%	-0.6693***	0.2005
			(0.2316)	(0.2064)
1	18	3%	-0.6693***	0.2005
			(0.2316)	(0.2064)
2	74	14%	-0.7135***	0.2150
			(0.2480)	(0.2095)
3	74	14%	-0.7135***	0.2150
			(0.2480)	(0.2095)
4	103	19%	-0.7355***	0.2023
			(0.2545)	(0.2105)
5	103	19%	-0.7355***	0.2023
			(0.2545)	(0.2105)
6	103	19%	-0.7355***	0.2023
			(0.2545)	(0.2105)
7	103	19%	-0.7355***	0.2023
			(0.2545)	(0.2105)
8	103	19%	-0.7355***	0.2023
			(0.2545)	(0.2105)

Notes: The table reports panel difference-in-discontinuities estimates of the headline specification after removing grid-connected control counties, at the county-ballot-year level. A control county lies in a state whose ballot failed. It is removed when a utility serving it also serves a *different* state that passed a ballot within the stated horizon of this county's own ballot. A horizon of zero counts only votes held in the same year, and eight counts the whole event window, so a larger horizon is a stricter check. All twelve failing ballots survive at every horizon, so no row reflects a change in which ballots are compared. The largest gap between any pair of grid-connected ballots in the sample is four years, which is why horizons above four remove nothing further. The specification follows Table C4, with county-by-ballot and calendar-year fixed effects, the estimation bandwidth $|R_{sa}| \leq 0.10$, and an event window of plus or minus eight years around the ballot. 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 C23. Interconnection-queue response to cannabis cultivation: solar and gas generators

Outcome	Queue applications		Withdrawals	
	(1)	(2)	(3)	(4)
Variables				
$\hat{\beta}$: Pass×Post	0.2207	0.2621	-0.1835	-0.1736
	(0.2563)	(0.2584)	(0.1296)	(0.1221)
Total-grid cultivation×Pass×Post	-0.0415***		0.0116	
	(0.0123)		(0.0145)	
Own-county cultivation×Pass×Post		-0.1985**		-0.0190
		(0.0753)		(0.0440)
Rest-of-grid cultivation×Pass×Post		-0.0285**		0.0140
		(0.0128)		(0.0123)
Fixed effects				
County × ballot	Yes	Yes	Yes	Yes
Calendar year	Yes	Yes	Yes	Yes
Fit statistics				
Observations	8,576	8,576	8,576	8,576

Notes: The table reports panel difference-in-discontinuities estimates on the stacked ballots, at the county-year level. The outcomes are county-year counts of solar and gas generator interconnection-queue applications in Columns (1)-(2) and dated withdrawals in Columns (3)-(4), from the LBNL Queued Up records. Columns (1) and (3) interact Pass×Post with total cannabis cultivation on the county's utility grid, the sum of own-county and rest-of-utility cultivation in the same state. Columns (2) and (4) decompose that grid dose into own-county and leave-one-out rest-of-grid cultivation, both interacted with Pass×Post in one regression, so the own-county coefficient is net of the rest-of-grid effect. Cultivation is satellite-detected area in km^2. The sample is the cultivation-dose subsample of 8,576 county-years in 16 states from 2017 onward. 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. 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 C24. 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, 2017-2023. Column (1) reports mean employment per county-year in workers, which is the mean across county-year cells of QCEW annual average employment, 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. 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.

Table C25. 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%
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%

Continued on next page

(Table C25 continued)

Rank	Occupation (SOC)	Share	Occupation (SOC)	Share	Occupation (SOC)	Share
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, as published by OEWS. 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, and the direct agriculture production codes are not included in the sector total either.

Table C26. 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			
Year FE	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 with a missing wage. 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. Three, two, and one asterisks denote statistical significance at the 1, 5, and 10 percent levels, respectively.

Table C27. 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 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 with a missing wage 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 C28. 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

Notes: 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 sample covers 2013-2024 and the 18 ballots inside the bandwidth with pre- and post-ballot trade. The unit of analysis is the state-year. The estimation bandwidth is $|R_s| \leq 0.10$. 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.