

The Impact of Data Centers on Local Economies

Diana Van Patten

diana.vanpatten@yale.edu

Yale University <https://orcid.org/0009-0003-7455-3011>

Fernando Alvarez

University of Chicago

David Argente

Yale University

Joyce Chow

Yale University

Social Sciences - Article

Keywords: Data centers, artificial intelligence, digital infrastructure, local economic effects

Posted Date: September 29th, 2026

DOI: <https://doi.org/10.21203/rs.3.rs-11089326/v1>

License:   This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Additional Declarations: There is **NO** Competing Interest.

The Impact of Data Centers on Local Economies

Fernando Alvarez* David Argente† Joyce Chow‡ Diana Van Patten§

September 2026

Abstract

Digital services are often experienced as placeless, but the infrastructure that supports them is not. Cloud computing and artificial intelligence are driving a global data-center investment boom of unprecedented scale. Non-random site selection, however, makes the effects of this expansion on host economies and local resource demand difficult to establish. We link facility-level data to US county outcomes and exploit historical determinants of fiber connectivity, combined with foreign growth in large-scale facilities, to address this challenge. The estimates reveal broad gains in employment, business activity and income. In long-run estimates for 1995–2023, a 10% increase in annualized data-center revenue raises total employment by approximately 2.7% and payroll by 3.6%. Employment and establishment gains persist, whereas construction effects are strongest at shorter horizons. House prices and public-supply water withdrawals also increase, while electricity-price estimates are positive in later years. We also find suggestive evidence of higher air pollution. These findings show that the infrastructure supporting the digital economy reshapes local economic activity while increasing demands on housing and shared resource infrastructure. Understanding these local effects is central to assessing the consequences of the global computing boom.

Keywords: Data centers; artificial intelligence; digital infrastructure; local economic effects

*University of Chicago. Correspondence: falvare@uchicago.edu.

†Yale University. Correspondence: david.argente@yale.edu.

‡Yale University. Correspondence: joyce.chow@yale.edu.

§Yale University. Correspondence: diana.vanpatten@yale.edu.

1 Introduction

Behind the apparent intangibility of digital services lies a physical infrastructure that connects their production to local economies and resources. Cloud computing, artificial intelligence, streaming, payments and enterprise software rely on facilities that occupy land, connect to fiber networks and require a continuous, reliable supply of electricity. These facilities are at the center of an infrastructure investment boom of unprecedented historical scale (PwC, 2026). Global investment in these facilities, known as data centers, nearly doubled between 2022 and 2024, reaching \$500 billion USD (International Energy Agency, 2025a), exceeded global investment in oil supply in 2025 (International Energy Agency, 2025b), and is expected to reach \$4 trillion USD before 2030 (International Energy Agency, 2026).

The local consequences of this investment boom are less clear. A data center can look attractive to local governments because it brings a large capital project, construction activity, tax revenue, and specialized service activity. At the same time, data centers are energy-intensive and spatially concentrated. They may also use scarce public infrastructure, like electricity and water. The result is a practical policy question: what are the effects on host counties?

This question has become more urgent with the rise of AI and a broader shift in data-center investment toward large, scale-oriented facilities, whose land, power, and infrastructure requirements are substantially greater than those of traditional enterprise data centers. Electricity use illustrates the scale of this transformation. Data centers accounted for around 1.5 percent of global electricity consumption in 2024, or about 415 TWh, and the United States accounted for the largest share of global data-center electricity use (International Energy Agency, 2025a). In the United States, data-center electricity use rose from 58 TWh in 2014 to 176 TWh in 2023, reaching about 4.4 percent of national electricity consumption, with substantial uncertainty but further growth projected through 2028 (U.S. Department of Energy, 2024; Shehabi et al., 2024). These national totals mask highly localized effects. Because data centers are spatially clustered, their effects on labor markets, utility systems, house prices, and electricity prices are concentrated in the few counties that host them.

This paper studies whether counties more exposed to data-center growth experience changes in employment, establishments, payroll, income, house prices, electricity prices, and public-supply water withdrawals. We leverage georeferenced facility-level data-center activity from the 451 Research Datacenter Knowledge Base (DCKB), a proprietary S&P Global/451 Research database with site-level information on facilities, capacity, and annualized revenue. We measure local economic outcomes using County Business Patterns (CBP), IRS Statistics of Income by county, FHFA house prices, EIA electricity prices, and public-supply water withdrawals from U.S. Geological Survey data.

The main identification challenge is that data centers do not locate randomly. Firms choose locations with favorable power prices, grid capacity, land, tax policy, climate, disaster risk, and fiber connectivity. We aim to address this concern with a shift-share instrumental-variable design based on historical corridor geography. The design builds on a key requirement of large, scale-oriented fa-

cilities: high-capacity fiber backbone connectivity. Such connectivity requires two complementary ingredients: being close to the fiber backbone itself and being able to access it through one of its entry points. Installing long-haul fiber across many separate parcels is costly and time-consuming, so underground fiber has often followed continuous transportation corridors, such as existing railroad lines, where routes are already assembled and rights-of-way may be easier to secure (Durairajan et al., 2015). Yet proximity alone is not enough: a cable that simply passes through a county does not provide usable connectivity unless the data center can access the high-capacity fiber backbone at a splice point, point of presence, or interconnection facility (Tang et al., 2013; Flanigan et al., 2009; Giotsas et al., 2015).

We therefore construct one historical measure for each margin of data-center suitability: one that captures the baseline value of proximity to a feasible long-distance route and a second one that captures the additional value created when that route also offers a usable access point into the fiber network. To recover the baseline route margin without relying on endogenous realized rail placement, we measure proximity to transcontinental routes predicted from terrain and fixed terminal pairs shown on the 1855 Pacific railroad survey map (Warren et al., 1855). To recover the additional margin of having access to the high-capacity fiber backbone, we interact concentrated 1870 rail coverage with internal access-node orientation, measured by endpoints and junctions, which often coincide with access points into the fiber. This interaction isolates whether corridor coverage becomes more predictive where the historical network offered a potential point of access. Together, the two measures provide pre-digital proxies for the margins of connectivity that shape data-center suitability. Several placebo tests show little evidence that the historical measures predict early economic levels or changes in outcomes before modern data-center expansion, although some early housing placebos reject zero. To the best of our knowledge, this is a new county-level historical rail-access instrument. Moreover, we believe this to be the first historical rail or network-topology instrument designed to predict county-level data-center development.

We use these historical measures in two complementary empirical designs. In a pooled analysis across years, we combine the baseline route and additional access margins into a predetermined county exposure and interact it with leave-U.S.-out growth in scale-oriented facility openings, including wholesale, powered-shell, and hyperscale sites. The intuition behind this shift-share design is that historical exposure determines *where* this external industry expansion should matter, while foreign growth determines *when* it should matter. By using growth in scale-oriented facilities, the first stage is weighted toward the large, expandable projects that have become increasingly important with cloud computing and AI. The estimates reveal a broad expansion of local economic activity. Data-center development has positive effects on total employment, construction employment, establishments, payroll, tax returns, adjusted gross income, wages, and public-supply water withdrawals. House prices and electricity prices also have positive point estimates, although they are less precisely estimated.

We then use the same two historical margins separately in dynamic long-difference estimates

to trace how local effects evolve over time.¹ These estimates clarify the timing of the effects. Employment and establishment effects remain positive at longer horizons, construction effects are strongest at shorter horizons, house-price increases persist, and positive electricity-price effects and public-supply water withdrawal estimates emerge toward the end of the sample. Together, the evidence suggests that the data-center development associated with the global expansion of scale-oriented facilities stimulates local employment, business activity, and income, while also putting upward pressure on house prices, electricity prices, and public water withdrawals.

The documented long-run effects are large. In the 1995–2023 long-difference specification, a 10 percent increase in annualized data-center revenue implies increases of 2.7 percent in total employment, 6.0 percent in construction employment, 2.3 percent in establishments, 3.6 percent in payroll, and 0.5 percent in electricity prices, together with increases of 1.4 percent in house prices, 14.7 percent in property-tax revenue, and 1.7 percent in fine-particulate concentrations, suggesting effects on air pollution. At the latest available endpoints, the corresponding estimates imply increases of 1.1 percent in tax returns, 2.3 percent in adjusted gross income, and 1.9 percent in wages through 2022, and 3.1 percent in public-supply water withdrawals through 2020.

For comparison, we develop an alternative shift–share based on proximity to modern long-haul fiber nodes, the 1980 urban-college share, and rest-of-world data-center growth. The signs of the coefficients are the same as in our benchmark case, but the magnitudes are smaller, and closer to the OLS case. For instance, the estimates of the effect on employment and payroll are about an order of magnitude smaller than those from the scale-oriented baseline. This design is geared toward development in established population centers and smaller enterprise-oriented facilities. Overall, we prefer our baseline design because the alternative exposures are less plausibly exogenous. We therefore treat this specification as a descriptive benchmark rather than as the preferred causal design.

Related literature. The paper relates to work on digital infrastructure and local labor markets (Forman et al., 2012; Kolko, 2012; Akerman et al., 2015), large plant openings and local multipliers (Greenstone et al., 2010; Moretti, 2010), and place-based development (Busso et al., 2013; Glaeser and Gottlieb, 2008). The historical-corridor design relates to work using old transportation networks to isolate infrastructure exposure, like Atack et al. (2010) and Donaldson and Hornbeck (2016) to study the effects of the nineteenth-century U.S. railroad network, Faber (2014) to address nonrandom highway placement, and Banerjee et al. (2020) to isolate transportation access in China. We adapt this route-based logic to digital infrastructure.

Finally, by addressing the well-documented endogenous site selection of data centers (JLL, 2026; CBRE, 2026, 2025) and spanning a wide range of outcomes, we complement recent evidence on data-center energy use and selected local effects (Shehabi et al., 2016; Masanet et al., 2020; Shehabi et al., 2024; Priest, 2026; Yue and Zeng, 2026; Pitt et al., 2026), as well as modeling studies that imply potentially large health damages from associated emissions (Han et al., 2026; Muller, 2026).

¹For this design, at any given endpoint, the foreign shift is common across counties and therefore adds no cross-county variation.

2 Data and Industry Background

Our analysis combines facility-level information on data centers with county-level measures of employment, business activity, income, house prices, electricity prices, and public-supply water withdrawals. The unit of analysis is a U.S. county observed over time.

Data-center activity The data-center panel comes from the 451 Research Datacenter Knowledge Base (DCKB), a proprietary facility-level database maintained by S&P Global Market Intelligence. It contains facility coordinates, ownership, operating dates, capacity measures (like utility power), facility type (for instance, wholesale, hyperscale, retail, etc), and annualized revenue (S&P Global Market Intelligence, 2026). We map U.S. facilities to counties and aggregate them to the county-year level. The data include records of data centers globally.

County outcomes County Business Patterns provides total employment, establishments, annual payroll, and construction employment (U.S. Census Bureau, 2026).² County IRS Statistics of Income files provide tax returns, adjusted gross income, and wages and salaries (Internal Revenue Service, 2026). FHFA repeat-sales indexes provide county house prices. House prices enter in natural logs in the pooled panel and as $\log(HPI_{ct}) - \log(HPI_{c,1995})$ in the long differences. EIA Form 861 provides utility revenue, electricity sales, and service-territory information used to construct county average retail electricity prices (U.S. Energy Information Administration, 2026). Public-supply withdrawals come from the USGS.³ Each outcome is followed through the final year available: 2023 for CBP, FHFA, and EIA outcomes, 2022 for IRS outcomes, and 2020 for public water supply.

Historical corridor inputs The historical design combines the 1870 railroad graph in Donaldson and Hornbeck (2016) and Atack et al. (2010) with 1890 county polygons, a fixed equal-area county lattice, the U.S. War Department’s 1855 *Map of Routes for a Pacific Railroad*, and NOAA’s ETOPO 2022 bedrock-elevation surface.⁴

2.1 Industry background

A useful way to think about a data center is as a factory for computation. It combines land, structures, electrical and cooling equipment, servers, specialized labor, and network connectivity

²For construction, we use SIC 15 and NAICS 236.

³Such data is available as [Public Supply Water Use Reanalysis, version 2.0](#). We allocate HUC12 withdrawals to counties using geographic-area weights and measure total public-supply withdrawals in annual-average million gallons per day, transformed by IHS. Public supply comprises withdrawals by public and private suppliers serving at least 25 people or at least 15 service connections for at least 60 days. For the dynamic placebo pre-2000, we use the USGS [quinquennial county water-use compilations](#).

⁴U.S. War Department, *Map of Routes for a Pacific Railroad* (1855), Library of Congress, <https://www.loc.gov/item/gm70005365/>; NOAA National Centers for Environmental Information, ETOPO 2022 Global Relief Model, <https://www.ncei.noaa.gov/products/etopo-global-relief-model>. The preferred sample contains counties with observed 1870 population and boundaries stable by 1870.

to produce computing capacity. Expansion increasingly takes the form of wholesale, powered-shell, and hyperscale sites designed for large and expandable computing loads. These are also the facilities most likely to expand with AI, given their concentrated computing capacity, reliable power, and room to add capacity. Panel A of Figure 1 documents revenue growth, with revenue reaching over 20 billion dollars in 2024 in the US alone, and Panel B shows the concentration of facilities across U.S. states. Panels C–E show the expansion of scale-oriented facilities, especially from the mid-2010s onward. Appendix Figure K.2 reports capacity growth.⁵ For instance, openings of scale-oriented facilities rose from 34 in 2013 to 215 in 2023. In total, the data record 1,457 such facilities built between 1995 and 2023 across 54 countries.

Data-center location depends heavily on fiber access and latency, grid capacity, electricity prices, land, climate, water availability, permitting, taxes, and disaster risk. Once a site is selected, building the power, cooling, and network infrastructure needed to operate it can generate construction activity and place significant demands on local utility systems. Supplemental Appendix K, Figures K.1–K.3, provides additional industry and within-US geographic detail.

3 Historical network exposure

Constructing a plausibly exogenous local exposure share is challenging in this setting. The share must predict where data centers can expand, yet it cannot simply restate contemporary advantages—fiber hubs, skilled labor, energy conditions, available land, or expected growth—because those factors likely shape employment, income, housing, and electricity prices. Most obvious predictors of data-center location are therefore poor candidates for causal identification precisely because they can affect the outcomes directly. We instead look far back in time for geographic variation that remains economically connected to modern backbone access.

3.1 Economic logic: a route and a place to access it

The design rests on a simple observation. Long-haul fiber often follows railroads and other transportation corridors with assembled rights-of-way because doing so lowers installation costs and reduces the time and complexity of securing a continuous route (Hogendorn, 2011; Durairajan et al., 2015). Yet, a long-distance fiber cable passing through a location does not by itself provide usable connectivity: a data center can use it only where there is a physical connection point that allows data to enter and leave the network. These access points are often located where routes meet, terminate, or branch (Tang et al., 2013; Durairajan et al., 2015; Flanigan et al., 2009; Giotsas et al., 2015).

Therefore, being close to a long-distance transportation corridor is valuable for a data center because it lowers the cost of reaching the fiber backbone, and that value is greater when the nearby network also contains a potential access point into that fiber cable, which often coincides with

⁵We classify scale-oriented facilities as wholesale, powered shell, and hyperscale. We assign each facility’s currently reported annualized revenue to its reported build cohort and cumulate those facility amounts at the county-year level.

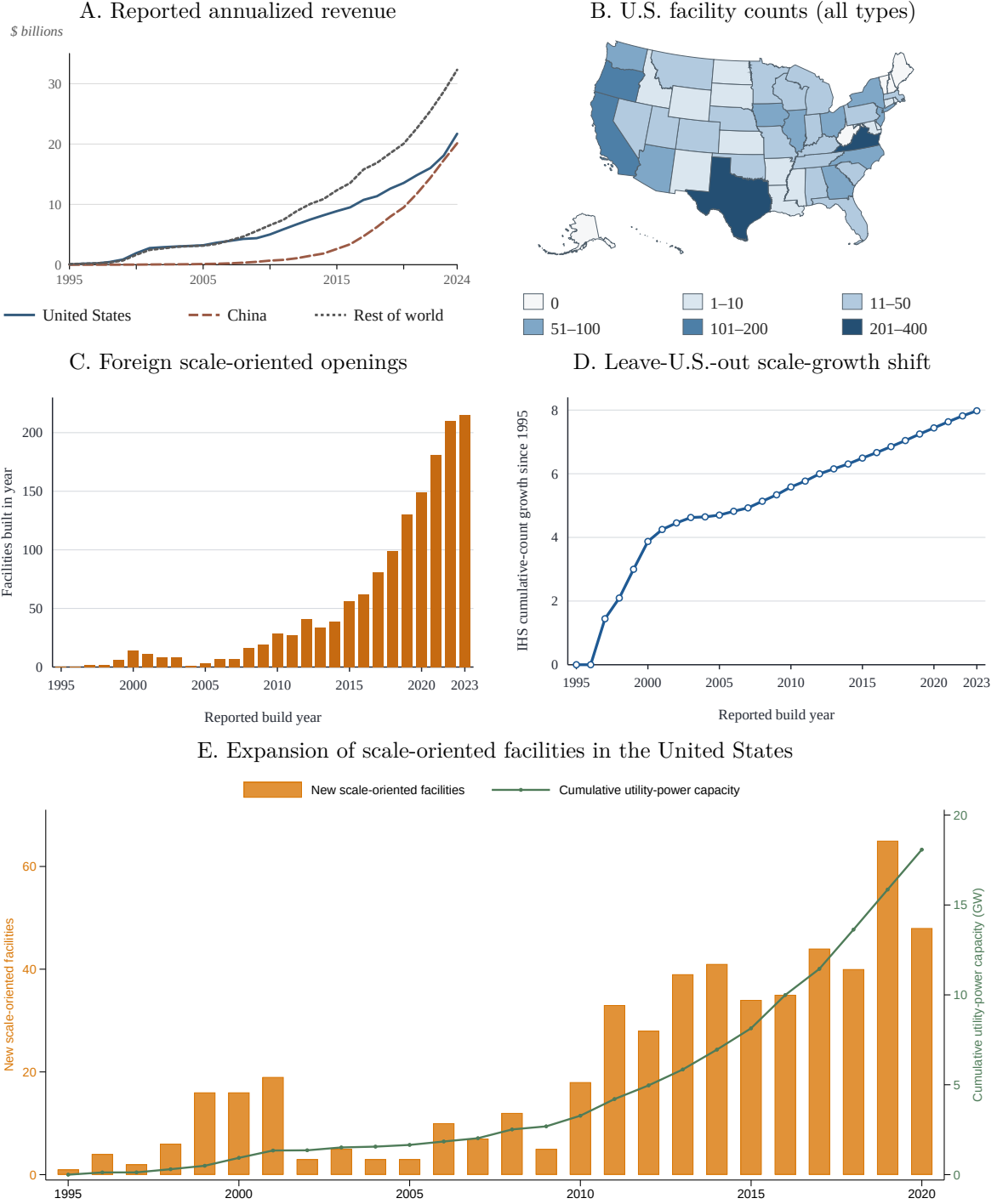


Figure 1: The expansion and geography of data centers

Notes: Panel A cumulates reported annualized revenue for all facility types, 1995–2024. Panel B maps 2,014 U.S. facility records of all types with reported build years in 1995–2025; shading gives the number of facilities in each state, including the District of Columbia. Zero means no qualifying record in this extract. Alaska and Hawaii appear as insets. Panels C–E concern scale-oriented facilities (Wholesale, Powered Shell, and Hyperscale sites). Panel C shows foreign annual openings and Panel D the cumulative-IHS growth shift G_t used in equation (2); both use 1,457 facilities built in 1995–2023 outside the United States. Panel E shows U.S. annual openings (bars) and cumulative utility-power capacity (line), 1995–2020. Source: 451 Research DCKB. State boundaries: U.S. Census Bureau.

points at which traffic can enter, leave, or branch from the long-haul route. This idea can be expressed as

$$\text{data-center suitability}_c = f(\underbrace{\text{route proximity}_c}_{\text{baseline corridor value}}, \underbrace{\text{access-node potential}_c}_{\text{additional value of local access}}). \quad (1)$$

The empirical challenge is that the two terms cannot be constructed credibly by simply inserting proximity to the realized railroad network into both of them. Railroads were built partly to connect existing population and economic activity, so realized rail proximity may remain associated with persistent determinants of the outcomes. We therefore construct two historical measures to recover each term in equation (1).

3.2 Recovering both margins from historical geography

Route proximity. For the baseline route-proximity term, we replace proximity to the realized railroad with proximity to terrain-predicted routes. Least-cost-route designs are useful because they recover the component of transport alignment driven by construction feasibility rather than the local population and economic activity that influenced realized placement. Prior work has used such routes to instrument for actual highway and railway connections when studying industrialization, structural change, and urban development in China, nineteenth-century England and Wales, and India (Faber, 2014; Bogart et al., 2022; Fenske et al., 2023). We adapt this route-based logic to digital infrastructure.

We begin with six east–west terminal pairs shown on, or tied to ports shown on, the 1855 Pacific railroad survey map (Warren et al., 1855). For each pair, the least-cost algorithm chooses the grid path with the lowest cumulative construction cost, where edge cost rises with distance and squared terrain grade. The resulting transcontinental network depends only on the fixed historical endpoints and physical terrain. The county measure increases with proximity to one of these predicted routes. This procedure is summarized in Figure 2, panel A. Neither realized railroad placement nor population, modern infrastructure, fiber, data centers, or economic outcomes enters the routing algorithm. Terrain-predicted proximity therefore captures baseline corridor suitability without relying on the realized location of railroads.

Local access. For the additional access-node term, the realized rail graph contains information that a hypothetical route cannot supply: whether a local corridor merely passed through or instead contained a place where lines met, ended, or branched. These points are valuable as they are likely to be access nodes into the fiber network (Flanigan et al., 2009). Therefore, we first compare the share of county land within 10 kilometers of the 1870 rail network with the share within 50 kilometers, distinguishing counties whose broad rail exposure is concentrated close to the line from those whose exposure is more diffuse, as shown in Figure 2, panel B.1. We then measure the relative prevalence of endpoints and junctions within the county. Namely, we compare the number of internal endpoints and junctions with the number of places where rail lines cross the county

boundary, as shown in Figure 2, panel B.2. After centering both measures, we interact them. A positive first-stage coefficient means that concentrated near-line coverage is more predictive of later data-center development when the historical network is more oriented toward an internal access node.

This interaction does not assume that every historical rail junction became a modern fiber node. Instead, it asks if having both concentrated rail coverage and access-node-oriented network geometry matters beyond the separate influence of either characteristic. The long-difference and balance specifications account separately for concentrated rail coverage and access-node orientation.⁶

In sum, both proposed instruments have distinct roles within one economic mechanism. Terrain-predicted route proximity substitutes for endogenous realized rail proximity in measuring the baseline value of being near a feasible long-distance corridor. The realized-rail interaction measures the additional value of proximity when the historical topology offered a local access point. The first supplies a model-predicted route; the second supplies access-node information available only in the realized route. Figure 2 summarizes the two constructions. Appendix A provides the exact construction of both measures.

This new instrument therefore combines local concentration of 1870 corridor coverage with orientation toward internal endpoints and junctions. While existing historical-rail designs motivate these ingredients, to the best of our knowledge, this measure is new. Furthermore, historical rail or network topology has not previously been used in the literature to instrument county-level data-center development.

3.3 A shift-share instrument

We combine the two historical measures into a single exposure share, denoted by s_c . Namely, we standardize the rail access-node interaction and terrain-route proximity, average them with equal weight, and rescale the result from zero to one. Appendix A gives the exact index and normalization. The resulting share translates equation (1) into predetermined county exposure. Its terrain component supplies the baseline route-proximity margin without using realized route placement as a standalone instrument. Its rail interaction supplies the additional access-node margin while absorbing the two lower-order rail measures. Large, expandable campuses may be particularly sensitive to both given their connectivity needs. Let G_t denote cumulative leave-U.S.-out growth in scale-oriented openings.⁷ Our shift-share is

$$Z_{ct} = s_c G_t. \quad (2)$$

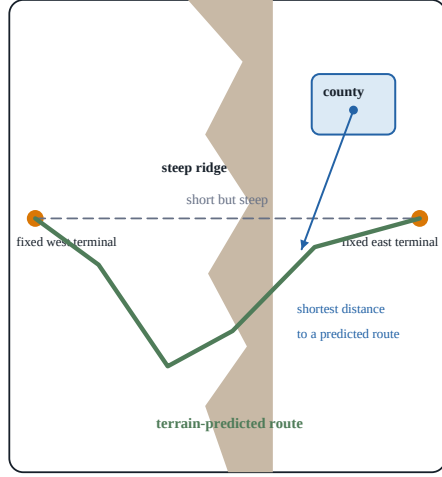
Identification follows two complementary views in the shift-share literature. [Borusyak et al.](#)

⁶In the pooled panel, county fixed effects absorb the time-invariant levels of these components. The centered interaction is used as a prespecified exposure component and is not interpreted as a monotone index of their underlying levels.

⁷No U.S. facility or county outcome enters the shift, so data-center growth in a particular U.S. county cannot mechanically generate it. The shift uses 1,457 scale-oriented facilities in 54 countries outside the United States; Appendix Figure A.1 shows the foreign shift, and Appendix A records the exact type classifications and sample rules.

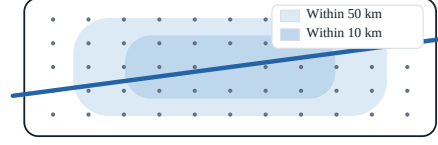
A. A route predicted from endpoints and terrain

Fix 1855 terminal pairs; let terrain determine the route



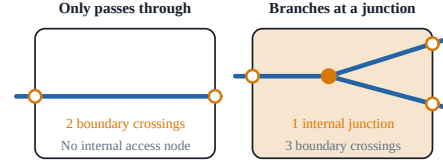
B. Rail proximity and junction access

1. Measure how much county land lies near the 1870 rail network



Near-versus-broad rail exposure: is coverage concentrated very close to the track rather than broadly nearby?

2. Compare internal endpoints and junctions with boundary crossings



Access-node geometry: internal endpoints and junctions relative to crossings

The rail measure combines proximity with access-node geometry.

Figure 2: Intuition for the two historical-network instruments

Notes: Panel A illustrates the baseline route margin: fixed endpoints and terrain generate a least-cost alignment whose proximity replaces endogenous proximity to the realized railroad as a standalone excluded variable. Panel B illustrates the additional access-node margin: concentrated exposure to a realized corridor is more valuable when the historical graph contains an internal endpoint or junction rather than merely passing through. The drawings are schematic and contain no coefficient or quantitative scale.

(2022) develop a quasi-experimental framework based on exogenous shocks. While our leave-U.S.-out measure captures an external expansion in scale-oriented facilities, Goldsmith-Pinkham et al. (2020) emphasize that identification can alternatively come from predetermined and conditionally exogenous exposure shares. We consider this to be the main source of identification in our design, since the county share is constructed entirely from nineteenth-century corridor geography and does not use modern data-center activity or county outcomes. The pre-expansion placebo tests reported in Section 5 further show that the historical allocation underlying the instrument does not systematically predict outcomes before the data-center expansion period.

4 Local economic activity expands: Pooled shift-share design

Pooled annual specification We begin our analysis with a shift-share specification that pools all years in our sample. We estimate

$$D_{ct} = \pi Z_{ct} + \alpha_c + \lambda_{s(c)t} + v_{ct}, \quad (3)$$

$$Y_{ct} = \beta \hat{D}_{ct} + \alpha_c + \lambda_{s(c)t} + \varepsilon_{ct}, \quad (4)$$

where D_{ct} is the IHS of reported annualized revenue attached to facilities built in county c by year t , Z_{ct} is defined in equation (2), α_c is a county fixed effect, and $\lambda_{s(c)t}$ is a state-by-year fixed effect. County fixed effects remove permanent county differences, while state-by-year fixed effects compare counties facing the same state-level conditions in each year. Outcomes use IHS except for logged house and electricity prices; standard errors are clustered by state. Because electricity prices are strictly positive, we use their logarithm. At positive revenue values, where IHS revenue behaves like a logarithm, its coefficient is an approximate elasticity. We report Anderson–Rubin inference alongside conventional 2SLS inference. The single coefficient summarizes the average persistent relationship between the accumulated data-center stock and the outcome over the sample.

The state-clustered first-stage Wald statistic exceeds the conventional $F = 10$ benchmark in eight of ten outcome samples, while the electricity-price and public-supply specifications remain weakly identified. For weak-IV inference, we report Anderson–Rubin p -values throughout and use them for the electricity-price and water-supply outcomes.

Pre-period validation The historical exposure is not systematically related to county economic conditions before modern data-center expansion. None of the 48 early-level balance tests rejects at the 5 percent level, and the pre-expansion placebo regressions show little evidence that the exposure predicts earlier changes in employment, establishments, payroll, income, house prices, or electricity prices. Results are similar when the rail access-node and terrain-route measures enter separately in the long-difference design. The complete early-level balance results are reported in Appendix F, Tables F.1–F.4. These patterns reduce concerns that the instrument captures counties that were already larger or following different economic trajectories.

Pooled results Figure 3 reports the pooled annual and latest-endpoint long-difference estimates, with implied effects at the mean sample change and at the median positive revenue increment evaluated at the mean positive-revenue base. Panel A of Supplementary Table E.3 reports the full pooled results. Counties whose data-center development is predicted to rise as scale-oriented facilities expand experience broad gains in employment, construction activity, business activity, and income; i.e., the corresponding estimated coefficients are positive and statistically significant. House-price and electricity-price coefficients are positive but imprecisely estimated. Total public-supply withdrawals are positive and statistically significant under weak-IV-robust inference.⁸

To grasp these magnitudes, consider the total-employment coefficient of 0.43; it implies that a 10 percent increase in data-center revenue raises employment by approximately 4.3 percent. Across the 1,577 counties in that estimation sample, the mean treatment change is about 24.3 percent.

⁸Appendix D reports the corresponding OLS estimates for both the pooled annual and latest-endpoint long-difference specifications. The fact that the OLS estimates are smaller than the IV estimates is consistent with negative selection in the location of data centers. Data centers may be disproportionately built in counties with cheap land, excess electricity capacity, or underused industrial sites. These characteristics make these locations attractive for investment but may also predict relatively slow economic growth in the absence of a data center. OLS would therefore compare data-center counties with places that have stronger underlying growth prospects, leading it to understate the positive causal effect of data-center investment. The latter is particularly likely for scale data centers as compared with traditional enterprise ones.

Evaluating the coefficient at that change gives an implied employment effect of 9.8 percent, as reported in column (1) of Figure 3. Column (2) of Figure 3 evaluates the median positive revenue increment at the mean positive-revenue base and implies 4.2 percent more employment.

Scale-oriented versus traditional data centers. Because our shift uses foreign growth in wholesale, powered-shell, and hyperscale facilities, its first stage is weighted toward large, expandable projects rather than smaller traditional sites. Scale-oriented facilities require more land and scalable power and connectivity, whereas third-party providers are more urban and buyer-proximate than major cloud providers, which concentrate in fewer, lower-density markets (Fang and Greenstein, 2025; Shehabi et al., 2016; CBRE, 2026). Thus, the estimates identify the effects of development induced by scale-oriented cloud and AI expansion, not the average effect of a data center—a key distinction for interpreting their magnitude. This is also the segment most likely to grow with AI (International Energy Agency, 2025a).

Appendix C develops a complementary design based on fiber-node proximity and the historical urban-college share, both interacted with the same rest-of-world revenue-growth shift. These exposures are more closely aligned with established network and population centers, and more tailored to capture the effects of traditional smaller-scale data centers. While their exogeneity claim is weaker, these measures provide a useful benchmark.⁹ On the common support, its total-employment coefficient is roughly one-tenth of the baseline estimate; most other point estimates are also smaller.

5 Effects persist over time: Dynamic historical-corridor design

Long-difference specification, controls, and inference The pooled specification estimates one coefficient for each outcome over the full annual panel. To understand the dynamics behind this average effect, the long-difference specification estimates a separate effect for each available endpoint from 2000 through 2023. At a given endpoint, foreign scale-oriented growth is common to all counties, so multiplying the historical exposure share by the shift would only rescale the share. Therefore, instead of the shift-share, we leverage the two underlying historical measures separately: the standardized rail access-node interaction, $\tilde{Z}_c^{AN}$, and terrain-route proximity, Z_c^{LC} .

For each endpoint t , we estimate

$$\Delta_t Y_c = \beta_t \widehat{\Delta_t D_c} + X'_c \delta_t + \alpha_s + \varepsilon_{ct}, \quad \Delta_t D_c \text{ instrumented by } (\tilde{Z}_c^{AN}, Z_c^{LC}), \quad (5)$$

where $\Delta_t Y_c = Y_{ct} - Y_{c,1995}$ and $\Delta_t D_c = D_{ct} - D_{c,1995}$. For house prices, $Y_{ct} = \log(HPI_{ct})$. The coefficient β_t measures the effect of cumulative data-center development on the change in the outcome between 1995 and endpoint t .

The “Early controls” specification includes historical population and geographic characteristics; it also separately controls for concentrated rail coverage and access-node orientation; and controls

⁹Appendix Tables C.2–C.9 report the alternative exposures’ early-level and pre-1995 placebo relationships.

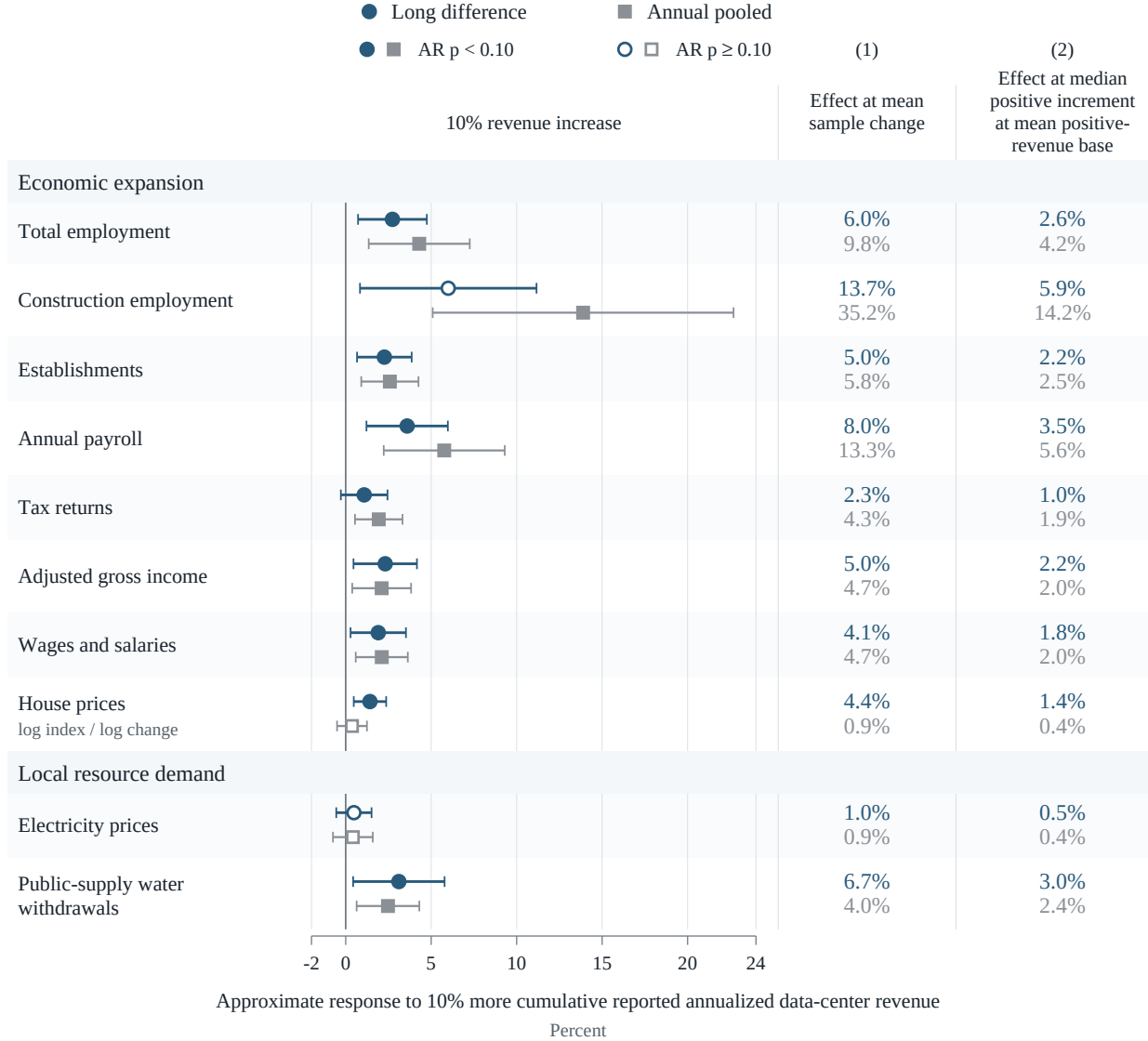


Figure 3: Economic expansion and local resource demand

Notes: Points give the approximate response to a 10 percent increase in cumulative data-center annualized revenue (10β). Blue circles and blue numbers show preferred long differences with Early + terrain controls; gray squares and gray numbers show pooled annual estimates with county and state-by-year fixed effects. Long differences begin in 1995 and end in 2023, except for tax returns, adjusted gross income, and wages (2022) and public-supply water withdrawals (2020). Pooled panels begin in 1995, except for water (2000), and use the same final years. The first numeric column evaluates the coefficient at the unweighted mean change in IHS cumulative revenue among the outcome's estimation counties. The second uses $\delta = \text{asinh}(110) - \text{asinh}(100)$: a rounded \$10 million increment at a \$100 million base motivated by the median positive county-by-entry-year increment and mean positive-revenue base. For a treatment change δ , these columns report $100[\exp(\beta\delta) - 1]$ percent. House prices enter in logs in the pooled model and log changes from 1995 in the long differences. All plotted responses are in percent. Percent interpretations are approximate at positive revenue values and, for IHS outcomes, at outcome values sufficiently far from zero. Bars are conventional 90 percent confidence intervals based on state-clustered standard errors. Marker symbols summarize Anderson–Rubin inference for a zero effect: ●, ■ for $p < 0.10$ and ○, □ for $p \geq 0.10$. The two numeric columns reproduce columns (6) and (7) of Supplementary Table E.3.

for economic levels measured in 1986 and 1989 and state fixed effects. The “Early + terrain” specification adds elevation, elevation dispersion, and slope. Appendix A provides the exact control list. Standard errors are clustered by state. When the first-stage statistic is below 10, inference and significance markers are based on the joint Anderson–Rubin weak identification test.

Appendix B reports the first-stage results in Table B.1 and Figure B.1. Both historical components become more predictive in their prespecified first-stage directions as cumulative data-center development grows. The appendix therefore reports Anderson–Rubin tests alongside every outcome and horizon.

Validation and long-difference placebos Figure 4 asks whether either corridor measure predicts county conditions that were already determined *before* modern data-center expansion. If the historical measures were just proxying for counties that were already larger or growing differently, they should predict these earlier economic levels. All outcomes are measured between 1985 and 1990. Panel A conditions on historical population and geography, and Panel B also includes direct terrain controls. The estimates are all statistically indistinguishable from zero: none of the 48 joint tests across the two historical cohorts and two control sets rejects at 5 percent.

The dynamic plots in Figure 5 and Appendix Figure E.1 add a second set of placebo checks using changes in the outcomes before 1995. These tests, in orange, show little evidence of employment, income, or electricity-price trends before data centers expanded. The housing placebos reject zero in 1987 at the 5 percent level and in 1986 at the 10 percent level under Early + terrain controls.¹⁰

5.1 Dynamic long-difference results

This section documents dynamic responses across outcomes. In each long-difference estimation, the post-period coefficient answers: by that end-year, how much more did the outcome change in counties whose accumulated data-center development was induced by the two historical corridor measures? The pre-period placebos ask whether the same geography predicted outcome changes before modern data-center expansion. The resulting figures therefore show when local responses emerge. Note that construction demand concentrated around facility build-out can produce a larger pooled estimate, whereas housing effects that accumulate over time can produce a larger final long-difference estimate.

Figure 3 and Panel B of Appendix Table E.3 summarize the estimates at the *latest* endpoint available for each outcome. Figure 5 and Appendix Figure E.1 report separately estimated annual endpoints from 2000 through each source’s final year, together with the available pre-period placebos. In these figures, we report both 90 percent 2SLS confidence intervals (shown by vertical bars)

¹⁰Because data-center revenue is not defined before 1995, these placebo regressions hold the treatment fixed at its 1995–2020 change and ask whether the instruments predict earlier movements in the outcome. They use the stated Early controls, omitting an early level only when that same level is the placebo outcome. The 1985 and 1990 public-supply placebo estimates are also centered near zero under both control sets, although they come from the older county compilations.

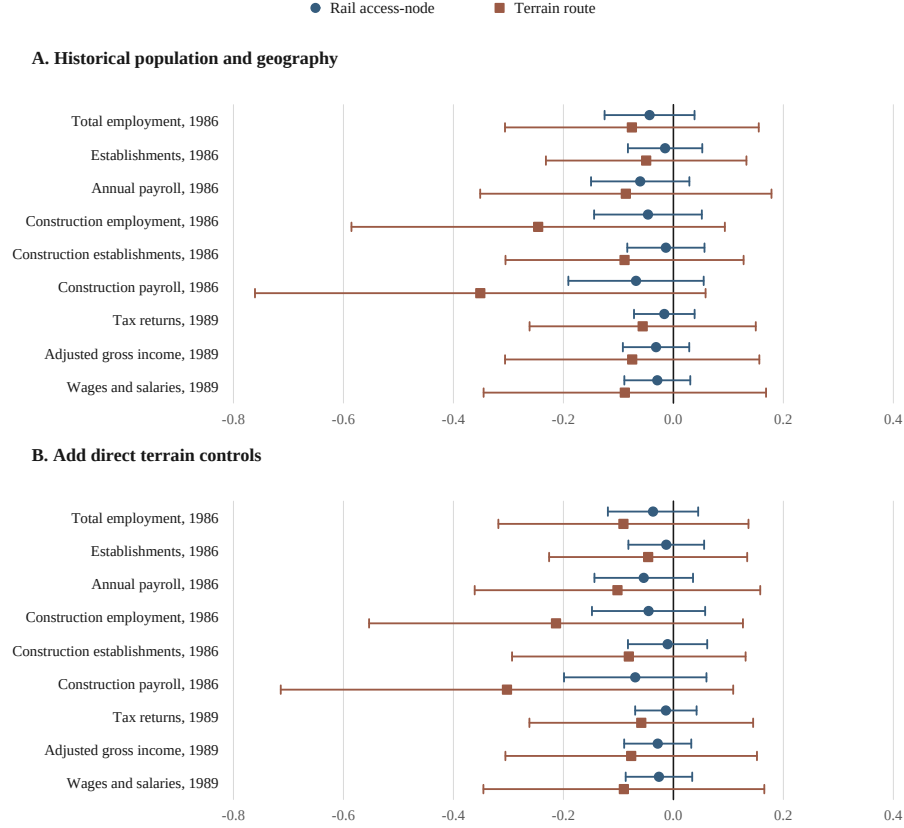


Figure 4: Balance on early county outcomes

Notes: Each point is the coefficient on one standardized historical-corridor measure when both instruments enter the same balance regression. Interval bars are 90 percent confidence intervals based on state-clustered standard errors. The preferred sample uses the 1870 cohort. Appendix F reports all joint tests.

and Anderson–Rubin inference; ●, ■ indicate AR $p < 0.10$, while ○, □ indicate AR $p \geq 0.10$.¹¹

Three remarks are in order. First, the pre-period estimates are generally near zero, with the housing exceptions noted above. Second, post-period intervals are wide, especially before the early 2010s. Precision then improves as the first stage strengthens, coinciding with the hyperscale expansion documented in Figure 1. Third, later estimates are consistently positive across the outcomes.

To grasp the magnitudes of Panel B of Appendix Table E.3, consider the 2023 total-employment coefficient of 0.274. This implies that 10 percent more data-center revenue leads to approximately 2.7 percent more employment between 1995 and 2023. As the mean treatment change is 23.8 percent, the implied employment effect is 6.0 percent, as reported in column (6). Column (7) evaluates the rounded median-increment benchmark at the mean positive-revenue base and implies 2.6 percent more employment. Figure 5 and Appendix Figure E.1 use the same coefficient units. The 2023 total-employment point is 0.259 with Early controls and 0.274 with Early + terrain

¹¹CBP, FHFA, and electricity-price outcomes extend through 2023, while IRS and water supply outcomes extend through 2022 and 2020, respectively.

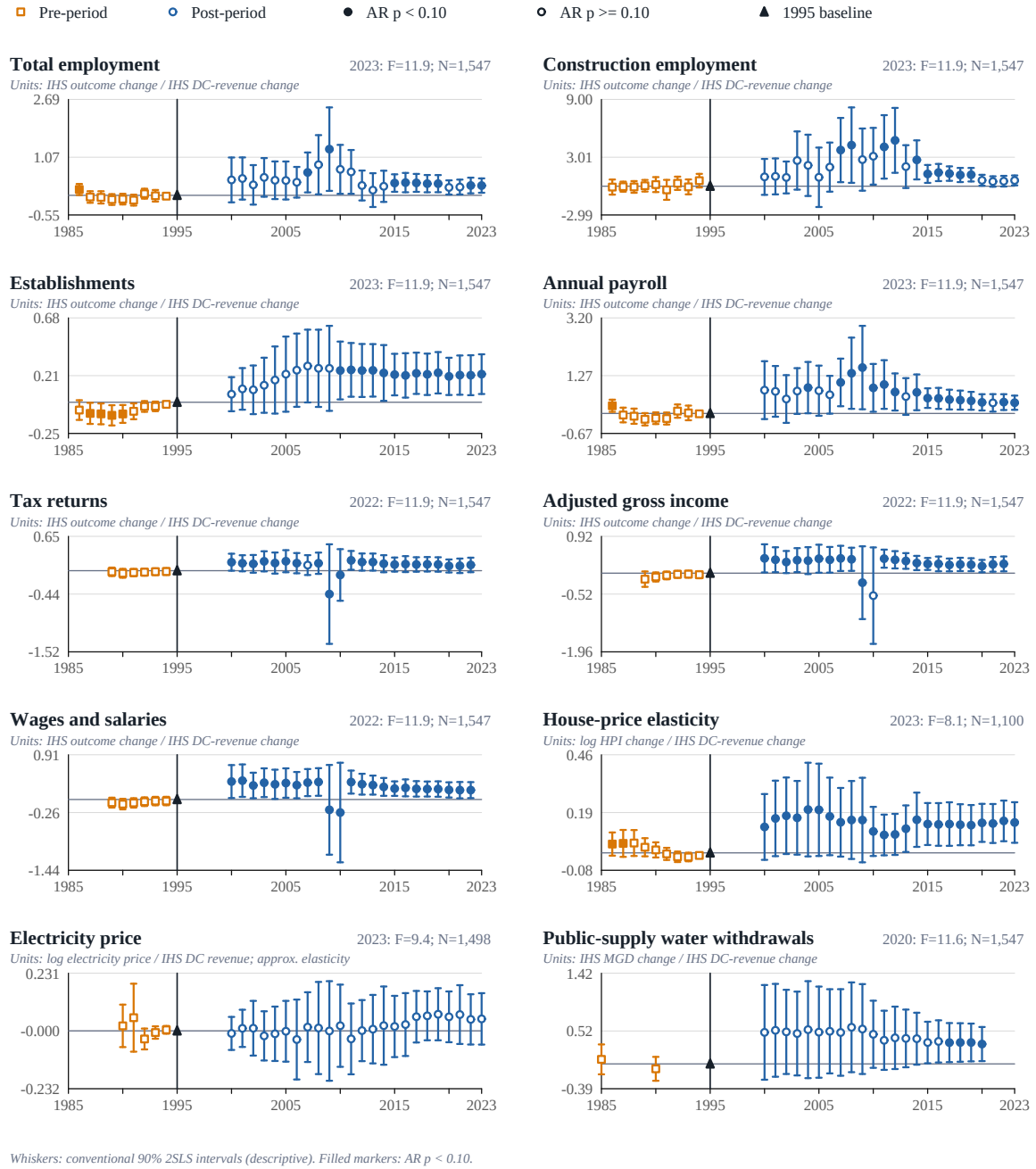


Figure 5: Annual historical-corridor outcome profiles: Early + terrain controls

Notes: Blue points are separately estimated post-1995 endpoints and orange points are separate pre-period placebo regressions. Interval bars show conventional 90 percent 2SLS confidence intervals. Each outcome ends in its final source-supported year; no endpoint is interpolated. ●, ■: Anderson–Rubin $p < 0.10$; ○, □: Anderson–Rubin $p \geq 0.10$. For IHS outcomes the vertical axis has the same approximate elasticity interpretation at positive values; house and electricity prices use log differences and therefore have an approximate elasticity interpretation at positive revenue values. In the water panel, orange points and the 1995–2000 bridge use the quinquennial county compilations, while annual changes after 2000 use the HUC12 reanalysis.

controls. Evaluated at the same mean revenue change of 23.8 percent, these coefficients imply employment effects of 5.7 and 6.0 percent, respectively.¹² We now briefly discuss each outcome.

Employment and business activity. Total employment is positive throughout the post-period profile and remains positive through 2023. Construction employment is also positive, with the largest coefficients earlier in the expansion and a smaller later effect, consistent with construction demand being strongest while facilities are built. Establishments and annual payroll are positive and remain statistically significant through 2023.

Income. Adjusted gross income and wages are mostly positive throughout the sample and remain positive through 2022. Tax returns are also generally positive, although less precise. The 2009–2010 coefficients are unusually noisy and sometimes negative. Their timing overlaps with the financial crisis and its immediate aftermath.

House and electricity prices. The log house-price effects are positive throughout the post-period profile. The log electricity-price coefficients are also positive in the later years, although they are imprecisely estimated under weak-IV-robust inference.¹³

Public-supply withdrawals Total public-supply withdrawals are positive at later endpoints. For instance, in Figure 5, the 2020 coefficient is 0.311. Thus, a 10 percent increase in data-center revenue implies approximately 3.1 percent more public-supply withdrawals; evaluated at the mean revenue increase in the sample, the implied increase is 6.7 percent.

Air pollution. Appendix G examines fine particulate matter (PM_{2.5}) and ground-level ozone. The 1995–2023 long-difference estimate implies approximately 1.7 percent higher PM_{2.5} concentrations for a 10 percent increase in cumulative annualized data-center revenue. Pooled estimates also suggest higher ozone concentrations, although its long-difference results are less conclusive.

Property-tax revenue. Additionally, Appendix H explores whether data-center development affects property-tax revenue in hosting counties. Coefficients are generally positive but imprecisely estimated across specifications.¹⁴ Under early controls, estimates become significant at the 10 percent level toward the end of the sample. For instance, at positive values, the preferred-IV estimate implies that a 10 percent increase in annualized data-center revenue is associated with approximately 15 percent greater property-tax-revenue growth between 1995 and 2023.

¹²For context, the sample mean dependent-variable change is 0.134 IHS points, approximately 14.3 percent employment growth.

¹³For electricity prices in 2020, the first-stage F is 9.10 and the Anderson–Rubin p -value is 0.185. At the 2023 endpoint, $F = 9.39$ and the AR p -value is 0.606.

¹⁴One reason for the limited precision may be that in some cases facilities have property-tax abatements (Gargano and Giacoletti, 2025; Bommarito and Bommarito, 2026).

This large effect has potentially important implications. West Des Moines is a case-study that illustrates how an initial data-center investment can propagate through local fiscal capacity. Microsoft first built a series of data-center campuses in the city. Property-tax revenue generated by those projects was then used to finance roads, sewer, water, and fiber infrastructure extending beyond the campuses; by FY2025, city records attributed about \$285 million in public infrastructure to data-center tax increment (City of West Des Moines, 2016, 2025; City of West Des Moines Community and Economic Development Department, 2025). The city subsequently built an open-access conduit network that lowered entry costs for broadband providers: Google Fiber became its first tenant and citywide provider, and five providers were leasing space by FY2025 (City of West Des Moines, 2020, 2025). The sequence provides a potential rationale for large local effects; data-center investment expanded fiscal capacity, shared infrastructure, and entry by other firms. This mechanism is related to the *Henry George* logic: infrastructure financed by an initial expansion of the tax base can raise surrounding land values and make additional land profitable to develop, potentially broadening the future tax base (Arnott and Stiglitz, 1979).

Spatial-equilibrium interpretation. We view data centers as a local export industry that raises demand for labor, land, utilities, and local services. A spatial-equilibrium model therefore predicts higher wages, in-migration, establishment entry, house prices, and utility demand (Roback, 1982; Moretti, 2010); shared infrastructure can further lower other firms’ entry costs (Kline and Moretti, 2014; Suárez Serrato and Zidar, 2016). This framework rationalizes the joint outcome pattern, while the magnitudes of the responses suggest that fiscal-infrastructure spillovers or other local multipliers may provide substantial amplification beyond the direct-demand channel. We leave the analysis of these considerations for future research. Appendix J formalizes the model and contrasts technology- and subsidy-driven entry.

Data-center-driven growth. Appendix I provides three checks supporting a data-center-specific interpretation. Before a county’s first mapped data center—often well after the initial internet boom—the annual instrument is not statistically related to any of 21 employment, establishment, or payroll outcomes across six non-hosting digital sectors. All ten outcome estimates remain positive in both designs after flexibly controlling for the county’s 1995 digital economy. Moreover, excluding counties with PeeringDB-listed interconnection facilities unmatched to our data makes all ten pooled and seven of ten preferred long-difference estimates larger; weak-IV-robust Anderson–Rubin tests reject zero for eight outcomes in each design.

6 Conclusion

Global investment in data centers and their supporting energy infrastructure is projected to approach US\$4 trillion between 2026 and 2030 (International Energy Agency, 2026). This investment wave is not only transforming computing; it is creating large, place-based shocks to the economies that supply land, power, water, and connectivity. Using pre-digital variation in the geography of

backbone connectivity, we show that data-center expansion induced by the global growth of large, scale-oriented facilities generates broad and persistent local economic effects. In the long-run estimates, a 10 percent increase in annualized data-center revenue raises total employment by 2.7 percent, establishments by 2.3 percent, and payroll by 3.6 percent. Employment and establishment gains persist, while construction effects are strongest earlier in the expansion. House prices and public-supply water withdrawals also rise, and electricity-price effects become positive at later horizons.

The breadth of these responses suggests that the local consequences of data-center investment extend well beyond the facilities themselves. The West Des Moines experience illustrates one possible amplification channel: data-center tax revenue financed roads, sewer, water, and fiber infrastructure, including an open-access conduit network that subsequently lowered entry costs for other providers. More generally, an initial data-center investment can expand local demand and fiscal capacity while also increasing pressure on housing and shared infrastructure. Our reduced-form estimates capture this combined local-equilibrium response rather than any one mechanism in isolation.

As artificial intelligence accelerates the build-out of large-scale computing capacity, the relevant economic question is therefore broader than how much computation these facilities produce. It is how one of the largest infrastructure investment waves of the digital era transforms the places that host it. The economic consequences of AI will be shaped not only by what computing enables, but also by what its physical infrastructure does to local economies.

References

- Akerman, A., Gaarder, I., Mogstad, M., 2015. The skill complementarity of broadband internet. *Quarterly Journal of Economics* 130, 1781–1824.
- Arnott, R.J., Stiglitz, J.E., 1979. Aggregate land rents, expenditure on public goods, and optimal city size. *The Quarterly Journal of Economics* 93, 471–500. URL: <https://doi.org/10.2307/1884466>, doi:10.2307/1884466.
- Atack, J., Bateman, F., Haines, M., Margo, R.A., 2010. Did railroads induce or follow economic growth? urbanization and population growth in the american midwest, 1850–1860. *Social Science History* 34, 171–197. doi:10.1017/S0145553200011202.
- Banerjee, A., Duflo, E., Qian, N., 2020. On the road: Access to transportation infrastructure and economic growth in china. *Journal of Development Economics* 145, 102442. doi:10.1016/j.jdeveco.2020.102442.
- Bogart, D., You, X., Alvarez-Palau, E.J., Satchell, M., Shaw-Taylor, L., 2022. Railways, divergence, and structural change in 19th century england and wales. *Journal of Urban Economics* 128, 103390. doi:10.1016/j.jue.2021.103390.
- Bommarito, J., Bommarito, M.J., 2026. Data Center Policy across the 50 States: A Survey of Incentives, Land Use, and Energy Regulation. SSRN Working Paper. SSRN. URL: <https://ssrn.com/abstract=6964278>, doi:10.2139/ssrn.6964278. posted July 14, 2026. Available at SSRN: <https://ssrn.com/abstract=6964278>.

480 Borusyak, K., Hull, P., Jaravel, X., 2022. Quasi-experimental shift-share research designs. *Review*
481 *of Economic Studies* 89, 181–213.

482 Busso, M., Gregory, J., Kline, P., 2013. Assessing the incidence and efficiency of a prominent place
483 based policy. *American Economic Review* 103, 897–947.

484 CBRE, 2025. North america data center trends h2 2025. URL: [https://www.cbre.com/insights/](https://www.cbre.com/insights/books/north-america-data-center-trends-h2-2025)
485 [books/north-america-data-center-trends-h2-2025](https://www.cbre.com/insights/books/north-america-data-center-trends-h2-2025). cBRE Research.

486 CBRE, 2026. Key factors to consider for effective data centre
487 site selection. URL: [https://www.cbre.co.uk/insights/articles/](https://www.cbre.co.uk/insights/articles/key-factors-to-consider-for-effective-data-centre-site-selection)
488 [key-factors-to-consider-for-effective-data-centre-site-selection](https://www.cbre.co.uk/insights/articles/key-factors-to-consider-for-effective-data-centre-site-selection). cBRE Research.

489 City of West Des Moines, 2016. Microsoft to build \$1.5–\$2 billion data center in west des moines.
490 *City News and Updates*. City of West Des Moines, Iowa.

491 City of West Des Moines, 2020. WDM to bring high-speed internet choices to residents and busi-
492 nesses with google fiber. *City News*. City of West Des Moines, Iowa.

493 City of West Des Moines, 2025. Annual Comprehensive Financial Report: Fiscal Year Ended June
494 30, 2025. Technical Report. City of West Des Moines, Iowa.

495 City of West Des Moines Community and Economic Development Department, 2025. 2025 Annual
496 Report and Development Retrospective. Technical Report. City of West Des Moines, Iowa.

497 Donaldson, D., Hornbeck, R., 2016. Railroads and american economic growth: A market access
498 approach. *Quarterly Journal of Economics* 131, 799–858. doi:[10.1093/qje/qjw002](https://doi.org/10.1093/qje/qjw002).

499 Durairajan, R., Barford, P., Sommers, J., Willinger, W., 2015. Intertubes: A study of the US
500 long-haul fiber-optic infrastructure, in: *Proceedings of ACM SIGCOMM*, pp. 565–578.

501 Faber, B., 2014. Trade integration, market size, and industrialization: Evidence from china’s na-
502 tional trunk highway system. *Review of Economic Studies* 81, 1046–1070. doi:[10.1093/restud/](https://doi.org/10.1093/restud/rdu010)
503 [rdu010](https://doi.org/10.1093/restud/rdu010).

504 Fang, T.P., Greenstein, S., 2025. Where the cloud rests: The economic geography of data centers.
505 *Strategy Science* 10, 404–420.

506 Fenske, J., Kala, N., Wei, J., 2023. Railways and cities in india. *Journal of Development Economics*
507 161, 103038. doi:[10.1016/j.jdeveco.2022.103038](https://doi.org/10.1016/j.jdeveco.2022.103038).

508 Flanigan, E., Howard, M., Volz, M., 2009. Rural Interstate Corridor Communications Study:
509 Report to States. Technical Report FHWA-HOP-09-021. Federal Highway Administration, Office
510 of Operations. URL: <https://ops.fhwa.dot.gov/publications/fhwahop09021/>.

511 Forman, C., Goldfarb, A., Greenstein, S., 2012. The internet and local wages: A puzzle. *American*
512 *Economic Review* 102, 556–575.

513 Gargano, A., Giacoletti, M., 2025. Subsidizing the Cloud: U.S. State Incentives to Data Centers.
514 SSRN Working Paper. SSRN. URL: <https://ssrn.com/abstract=5881105>, doi:[10.2139/ssrn.](https://doi.org/10.2139/ssrn.5881105)
515 [5881105](https://doi.org/10.2139/ssrn.5881105). first version December 7, 2025; revised June 30, 2026.

- Giotsas, V., Smaragdakis, G., Huffaker, B., Luckie, M., Claffy, K.C., 2015. Mapping peering interconnections to a facility, in: Proceedings of the 11th ACM Conference on Emerging Networking Experiments and Technologies, pp. 37:1–37:13. doi:[10.1145/2716281.2836122](https://doi.org/10.1145/2716281.2836122).
- Glaeser, E.L., Gottlieb, J.D., 2008. The economics of place-making policies. *Brookings Papers on Economic Activity* , 155–253.
- Goldsmith-Pinkham, P., Sorkin, I., Swift, H., 2020. Bartik instruments: What, when, why, and how. *American Economic Review* 110, 2586–2624.
- Greenstone, M., Hornbeck, R., Moretti, E., 2010. Identifying agglomeration spillovers: Evidence from winners and losers of large plant openings. *Journal of Political Economy* 118, 536–598.
- Han, Y., Wu, Z., Li, P., Wierman, A., Ren, S., 2026. The unpaid toll: Estimating and addressing the public health impact of data centers. URL: <https://arxiv.org/abs/2412.06288>, [arXiv:2412.06288](https://arxiv.org/abs/2412.06288). arXiv preprint arXiv:2412.06288.
- Hogendorn, C., 2011. Excessive(?) entry of national telecom networks, 1990–2001. *Telecommunications Policy* 35, 920–932. doi:[10.1016/j.telpol.2011.09.003](https://doi.org/10.1016/j.telpol.2011.09.003).
- Internal Revenue Service, 2026. SOI tax stats: County data. *Statistics of Income county income data*.
- International Energy Agency, 2025a. Energy and AI. Technical Report. International Energy Agency. Paris. URL: <https://www.iea.org/reports/energy-and-ai>.
- International Energy Agency, 2025b. World Energy Outlook 2025. Technical Report. International Energy Agency. Paris. URL: <https://www.iea.org/reports/world-energy-outlook-2025>.
- International Energy Agency, 2026. Key Questions on Energy and AI. World Energy Outlook Special Report. International Energy Agency. Paris. URL: <https://www.iea.org/reports/key-questions-on-energy-and-ai>.
- JLL, 2026. 2026 global data center outlook. URL: <https://www.jll.com/en-us/insights/market-outlook/data-center-outlook>. jones Lang LaSalle Research.
- Kline, P., Moretti, E., 2014. People, places, and public policy: Some simple welfare economics of local economic development programs. *Annual Review of Economics* 6, 629–662. doi:[10.1146/annurev-economics-080213-041024](https://doi.org/10.1146/annurev-economics-080213-041024).
- Kolko, J., 2012. Broadband and local growth. *Journal of Urban Economics* 71, 100–113.
- Masanet, E., Shehabi, A., Lei, N., Smith, S., Koomey, J., 2020. Recalibrating global data center energy-use estimates. *Science* 367, 984–986.
- Moretti, E., 2010. Local multipliers. *American Economic Review* 100, 373–377. doi:[10.1257/aer.100.2.373](https://doi.org/10.1257/aer.100.2.373).
- Muller, N.Z., 2026. Measuring the Impact of Data Centers in the United States Economy: Monetary Damage from Air Pollution and Greenhouse Gas Emissions. Working Paper 35100. National Bureau of Economic Research. URL: <https://www.nber.org/papers/w35100>, doi:[10.3386/w35100](https://doi.org/10.3386/w35100).

- Pitt, D., Suen, I., Plisko, E., 2026. Localized Air Pollution Impacts from Data Centers in Northern Virginia. White Paper. Virginia Commonwealth University, Institute for Sustainable Energy and Environment. URL: https://scholarscompass.vcu.edu/isee_pubs/3/.
- Priest, A., 2026. Not in my back yard! the effects of data centers on housing prices. URL: <https://ssrn.com/abstract=6314620>, doi:10.2139/ssrn.6314620. SSRN Working Paper.
- PwC, 2026. Global Data Centre Outlook 2026–50: Where \$31.6 Trillion of Capex Flows in the Era-Defining AI Build-Out. Technical Report. PwC. URL: <https://www.pwc.com/gx/en/1/services/consulting/technology/data-centre-outlook.html>.
- Roback, J., 1982. Wages, rents, and the quality of life. *Journal of Political Economy* 90, 1257–1278. doi:10.1086/261120.
- Shehabi, A., Smith, S.J., Hubbard, A., Newkirk, A., Lei, N., Siddik, M.A.B., Holecek, B., Koomey, J.G., Masanet, E.R., Sartor, D.A., 2024. 2024 United States Data Center Energy Usage Report. Technical Report LBNL-2001637. Lawrence Berkeley National Laboratory.
- Shehabi, A., Smith, S.J., Sartor, D.A., Brown, R.E., Herrlin, M., Koomey, J.G., Masanet, E.R., Horner, N., Azevedo, I.L., Lintner, W., 2016. United States Data Center Energy Usage Report. Technical Report LBNL-1005775. Lawrence Berkeley National Laboratory.
- S&P Global Market Intelligence, 2026. 451 research datacenter knowledgebase. Product documentation and database accessed through Yale subscription.
- Suárez Serrato, J.C., Zidar, O., 2016. Who benefits from state corporate tax cuts? a local labor markets approach with heterogeneous firms. *American Economic Review* 106, 2582–2624. doi:10.1257/aer.20141702.
- Tang, X., Ghosh, S., Durairajan, R., Eriksson, B., Barford, P., 2013. Internet atlas: A geographic database of the internet, in: *Proceedings of the 5th ACM Workshop on HotPlanet*.
- U.S. Census Bureau, 2026. County business patterns. County-level employment, establishment, and annual payroll files.
- U.S. Department of Energy, 2024. DOE releases new report evaluating increase in electricity demand from data centers. Press release summarizing the 2024 United States Data Center Energy Usage Report.
- U.S. Energy Information Administration, 2026. Annual electric power industry report, form EIA-861 detailed data files. Electricity sales, revenue, customer, and service-territory data.
- Warren, G.K., Bien, J., Davis, J., United States War Department, 1855. Map of routes for a pacific railroad, compiled to accompany the report of the hon. jefferson davis, secretary of war. Library of Congress, Geography and Map Division. URL: <https://www.loc.gov/item/gm70005365/>.
- Yue, D., Zeng, Y., 2026. The local economic effects of data center entry. Working paper, SSRN.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [NATUREDATACENTERSAAVAPPENDIX.pdf](#)