Data Centers and Local Economies in the Age of AI: A Shift–Share Approach

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

September 2026

Abstract

Data centers are the physical infrastructure behind cloud computing, artificial intelligence, and enterprise software. The rapid diffusion of artificial intelligence is increasing demand for computing capacity, accelerating investment in data centers, and raising concerns about their local economic and environmental effects. We assemble a facility-level panel of U.S. data centers and link it to county-level measures of employment, establishments, payroll, income, house prices, electricity prices, and public-supply water withdrawals. To address endogenous site selection, we develop a shift-share instrument motivated by two complementary requirements of fiber backbone connectivity: a feasible fiber route and a point of access into that fiber. Because installing long-haul fiber across many separate parcels is costly and time-consuming, providers have often followed continuous transportation corridors, such as existing railroad lines, where routes are already assembled and access or rights-of-way are less burdensome. At the same time, a data center must be able to enter the backbone through a splice or access point. We construct two historical exposure measures. The first identifies locations where terrain made long-distance routes relatively inexpensive to build, while the second identifies access points into the fiber backbone. We interact these historical shares with growth in scale-oriented facilities outside the U.S. The resulting instrument does not systematically predict county outcomes before the modern expansion of data centers. After that expansion begins, the estimates show positive effects on total employment, construction employment, establishments, payroll, tax returns, adjusted gross income, and wages. We complement the shift-share design with dynamic long-difference specifications that use the same exogenous shares. The estimates show persistent gains in employment and establishments, construction effects that are stronger at shorter horizons, sustained increases in house prices and public-supply withdrawals, and positive effects on electricity prices that emerge later. Overall, this pattern is consistent with the hypothesis that data centers increase economic activity, both directly and indirectly, which puts pressure on house and electricity prices and increases water demand.

Keywords: Data centers; artificial intelligence; digital infrastructure; local economic effects; shift-share instruments; historical railroads.

*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

Digital services are often experienced as placeless, but the infrastructure that supports them is not. Behind cloud computing, artificial intelligence, streaming, payments, enterprise software, and other digital services are physical facilities that occupy land, connect to fiber networks, draw power continuously, and require grid reliability. These facilities, known as data centers, have become one of the most visible elements of the digital economy.

Data-center development can have substantial local effects. 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, 2025). 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 facilities: 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-fiber, 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 that the historical measures do *not* predict early economic levels or changes in outcomes before modern data-center expansion. 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 cumulative reported annualized data-center revenue leads to increases of 2.7 percent

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

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 an additional 4.4 percentage points of house-price growth. 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 sign of the coefficients are the same as in our benchmark case, but the magnitude 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. This paper is related to work on digital infrastructure and local labor markets. Broadband and internet infrastructure can affect wages, productivity, and firm location, but the effects depend on technology, adoption, and local absorptive capacity (Forman et al., 2012; Kolko, 2012; Akerman et al., 2015). Data centers, however, differ from household broadband access. They are spatially concentrated capital projects, and their effects operate through construction activity, specialized operations, electricity demand, tax base, and local supply chains. The paper also relates to research on large plant openings and place-based economic development. Greenstone et al. (2010) use runner-up locations for large plant openings to study agglomeration spillovers. Moretti (2010a) studies local multipliers, while Busso et al. (2013) and Glaeser and Gottlieb (2008) discuss the incidence and efficiency of place-based policies. Data centers are relevant in this context because local governments often compete for them, but their labor and power intensity differ from those of traditional manufacturing plants.

The historical-corridor design relates to work using old transportation networks or hypothetical routes to isolate infrastructure exposure from contemporaneous demand. Atack et al. (2010) and Donaldson and Hornbeck (2016) study the economic effects of the nineteenth-century U.S. railroad network. Faber (2014) uses least-cost spanning-tree routes to address nonrandom highway placement, while Banerjee et al. (2020) uses lines connecting historical cities to isolate transportation access in China. We adapt this route-based logic to digital infrastructure: the interaction of nineteenth-century corridor coverage with access-node orientation captures route-and-node complementarity, while terrain-only transcontinental routes capture where a long-distance connection could be built at relatively low geographic cost.

Finally, the paper relates to research on data-center electricity use. Shehabi et al. (2016) estimate U.S. data-center energy use, Masanet et al. (2020) reassess global data-center energy demand in light of efficiency improvements, and Shehabi et al. (2024) report recent U.S. data-center energy trends and future scenarios. We study a different margin: how the local arrival and expansion of data centers affects county employment, establishments, income, house prices, electricity prices, and public-supply water withdrawals. Contemporaneous work by Yue and Zeng (2026) studies the local effects of data-center entry using a difference-in-differences design. Entry, however, is likely endogenous because data centers are sited based on considerations like infrastructure, energy access, permitting conditions, tax incentives, expected growth, and other local advantages (e.g., JLL 2026; CBRE 2026, 2025). Our designs instead use predetermined historical corridor geometry, first interacted with foreign growth in scale-oriented facilities in a Bartik (1991) tradition and then unpacked into its two underlying components as instruments for cumulative county development.

The remainder of the paper is organized as follows. Sections 2 describes the data and industry background. Section 3 develops the historical-corridor measures and pooled shift–share design, and Section 4 presents the validation exercises and pooled estimates. Section 5 introduces the dynamic long-difference design, presents its validation and first-stage evidence, and reports the dynamic results. Section 6 concludes. The appendices provide construction details, additional first-stage and balance evidence, the alternative shift–share benchmark, OLS comparisons, and complete results.

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. We use the facility data to measure where data centers operate and how their scale evolves, and we use public county-level sources to measure the local economic conditions that may respond to this expansion.

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, powered shell, retail, etc), and currently reported annualized revenue (S&P Global Market Intelligence, 2026).[2] 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).[3] 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. 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). Each outcome is followed only through the final year available in its source: 2023 for CBP, FHFA, and EIA outcomes, 2022 for IRS outcomes, and 2020 for public water supply. Public-supply withdrawals come from the USGS, which provides annual

[2] We group facilities as: Scale-oriented: wholesale, powered shell, and hyperscale; more traditional/network-oriented: retail, telco, interconnect, hosting, cloud, reseller, small reseller, and system integrator. We assign each facility's currently reported annualized revenue to its reported build cohort and cumulate those facility amounts at the county-year level.

[3] For construction, we use SIC 15 in the SIC years and NAICS 236 in the NAICS years.

2000–2020 estimates at the HUC12 watershed level.[4] We allocate HUC12 withdrawals to counties using geographic-area weights and measure total public-supply withdrawals in annual-average million gallons per day (MGD), transformed by IHS.[5]

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.[6]

2.1 Industry background

Counts and geography. Data centers house servers, storage, and networking equipment together with the power and cooling systems needed to keep that equipment running. Demand has grown with cloud computing, software-as-a-service, streaming, digital public services, and, more recently, AI training and inference. Panel A of Figure 1 shows that the stock of facilities rises sharply after the mid-2000s in Asia-Pacific, the United States and Canada, and Europe. Panel B shows the concentration of U.S. facilities across states. Virginia is the largest market in the facility records, followed by Texas and California. 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 substantial construction activity and place significant demands on local utility systems.

Revenue and growth. 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 to produce computing capacity. Figure 2 summarizes the industry's expansion using four measures of facility scale: reported annualized revenue, operational space, UPS power capacity, and racks.[7] All four measures rise substantially, with growth accelerating from the mid-2010s onward. Magnitudes are also very large: by 2024, the reported annualized revenue in the U.S. is over 20 billion dollars and operational capacity surpasses 90 million square feet.

The rise of scale-oriented facilities. A growing share of data-center expansion has taken the form of wholesale, powered-shell, and hyperscale sites rather than traditional enterprise facilities built primarily for a single firm's internal computing needs, as evidenced in Figure 3 for the U.S. These scale-oriented facilities are designed to accommodate large and expandable computing loads, with

[4] Such data is available as Public Supply Water Use Reanalysis, version 2.0.

[5] 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 profile, the 1985 and 1990 points and the 1995–2000 bridge use the USGS quinquennial county water-use compilations; annual changes after 2000 use the HUC12 reanalysis.

[6] U.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.

[7] To construct the figure, we assign each facility's currently reported attributes to its build cohort and cumulate across cohorts. The series therefore measure the reported scale of facilities built by each year.

substantial power capacity, large campuses, and infrastructure that can grow as demand increases. They are also the facilities most likely to expand with AI, since large-scale training and inference require concentrated computing capacity, reliable power, and the ability to add capacity rapidly. Foreign openings of scale-oriented facilities accelerated sharply during the 2010s, rising from 34 in 2013 to 149 in 2020 and 215 in 2023. In total, the data record 1,457 such facilities built between 1995 and 2023 across 54 countries outside the United States. Figure A.1 documents this expansion.

3 Historical-corridor shift-share design

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

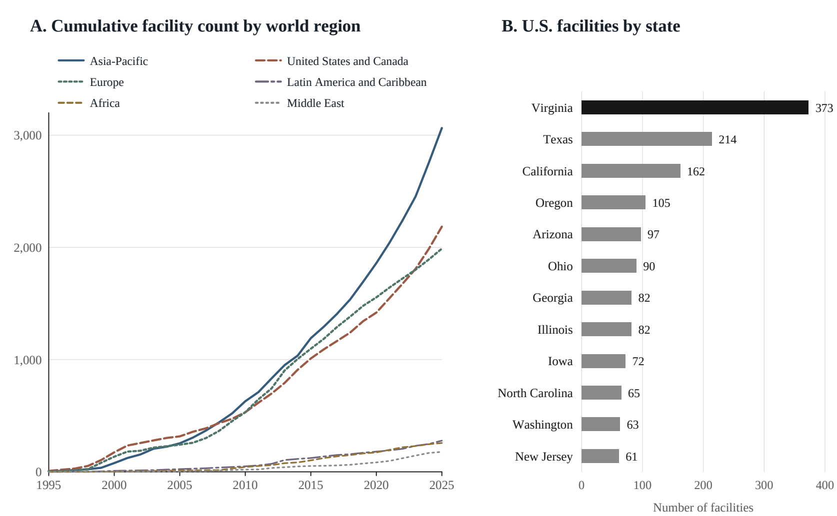

Figure 1: Data-center growth and geographic concentration

Notes: Panel A reports cumulative facility counts by subregion and reported build year from 1995 through 2025. Panel B reports the twelve U.S. states with the largest facility counts. Both panels use all records with a reported build year in the 451 Research DCKB. To read the scale, the first North American point in Panel A represents 8 facilities built by 1995; Virginia's bar in Panel B represents 373 facilities.

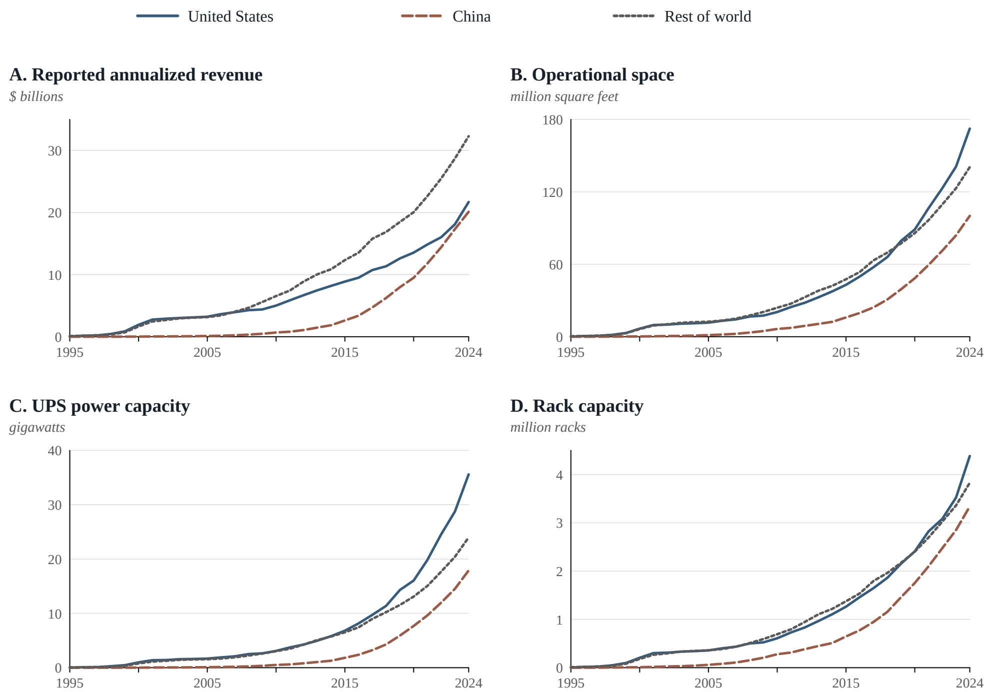

Figure 2: Reported scale of data-center facilities built by year

Notes: The panels cumulate facility attributes by reported build year from 1995 through 2024. Revenue is reported annualized revenue in billions of U.S. dollars; space is millions of square feet; UPS capacity is gigawatts; and racks are millions. To read the first U.S. observations in 1995, the plotted cohorts jointly report about $33.9 million in annualized revenue, 91,200 square feet, 10.5 megawatts of UPS capacity, and 3,247 racks. These are current reported attributes assigned to build cohorts, not realized historical flows.

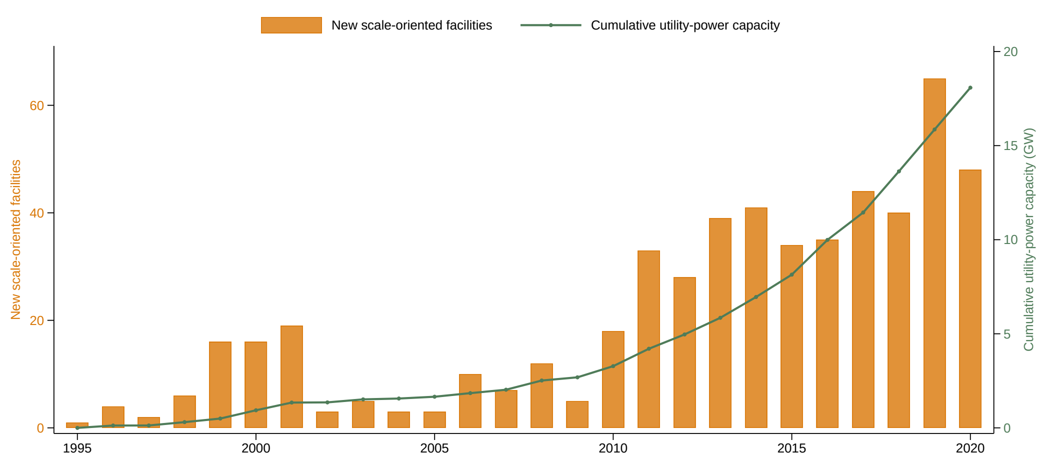

Figure 3: Expansion of scale-oriented data centers in the United States

Notes: Bars report the number of newly built scale-oriented facilities in each year on the left axis. Scale-oriented facilities are those classified as Wholesale, Powered Shell, or Hyperscale in the 451 Research Datacenter Knowledge Base. The line reports the cumulative reported utility-power capacity of these facilities, in gigawatts, on the right axis. Utility-power capacity measures the electrical capacity that the grid can deliver to a facility. Each facility's reported capacity is assigned to its construction year and then cumulated across construction cohorts.

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

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

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 us summarized in Figure 4, 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 4, 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 4, 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 whether 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. [8]

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 4 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.[9] Our shift-share is

$$Z_{ct} = s_c G_t. \tag{2}$$

Identification follows two complementary views in the shift–share literature. Borusyak et al. (2022) develop a quasi-experimental framework based on exogenous shocks. While our leave-U.S.-out measure

[8] In 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.

[9] No 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.

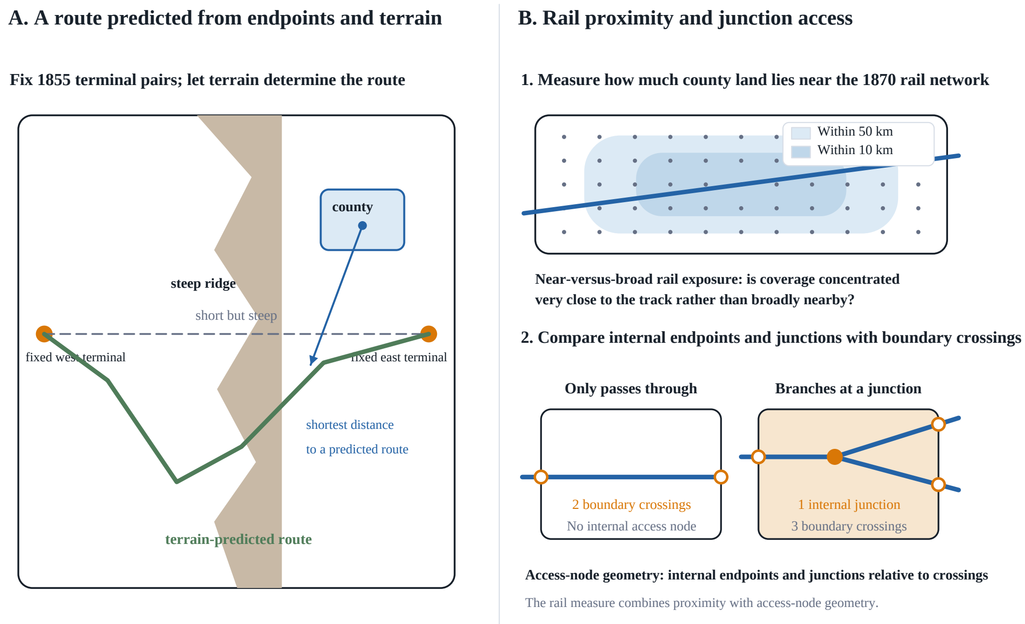

Figure 4: Intuition for the two historical-network instruments

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.

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 is 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 4 further show that the historical allocation underlying the instrument does not systematically predict outcomes before the data-center expansion period.

4 Pooled shift-share: specification and results

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

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

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

where D_{ct} is the IHS of cumulative 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 cumulative 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.

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 Table 1 reports one pooled annual estimate for each outcome. Counties whose data-center development is predicted to rise as scale-oriented facilities expand abroad 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 Anderson–Rubin inference.[10]

To grasp these magnitudes, consider the total-employment coefficient of 0.430, which implies that a 10 percent increase in cumulative data-center revenue raises employment by approximately 4.3 percent. Across the 1,577 counties in that estimation sample, the mean 1995–2023 treatment change is about 24.3 percent. Evaluating the coefficient at that change gives an implied employment effect of 9.8 percent, as reported in column (6) of Table 1. Column (7) 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. The pooled estimates may appear intuitively large. This aligns with shift being constructed from foreign growth in wholesale, powered-shell, and

[10]Appendix 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.

Table 1: Pooled historical-corridor shift-share estimates using scale-oriented foreign growth

Outcome	End	Estimate	(SE)	p-value	AR p	Implied effect at: Mean sample change	Implied effect at: Median increment at mean positive-revenue base
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Total employment	2023	0.430**	(0.180)	0.022	0.005	9.8%	4.2%
Construction employment	2023	1.389**	(0.535)	0.013	0.013	35.2%	14.2%
Establishments	2023	0.258**	(0.102)	0.015	0.003	5.8%	2.5%
Annual payroll	2023	0.576**	(0.215)	0.011	< 0.001	13.3%	5.6%
Tax returns	2022	0.193**	(0.084)	0.028	0.016	4.3%	1.9%
Adjusted gross income	2022	0.210*	(0.105)	0.052	0.028	4.7%	2.0%
Wages and salaries	2022	0.211**	(0.093)	0.028	0.008	4.7%	2.0%
House-price index	2023	0.037	(0.053)	0.492	0.478	0.9%	0.4%
Electricity-price elasticity	2023	0.042†	(0.071)	0.556	0.550	0.9%	0.4%
Water withdrawals	2020	0.247***†	(0.111)	0.033	0.009	4.0%	2.4%
Outcome sample	**Start**	**End**				**First-stage F**	
CBP outcomes	1995	2023				11.18	
IRS outcomes	1995	2022				10.80	
House prices	1995	2023				11.23	
Electricity prices	1995	2023				8.79†	
Water withdrawals	2000	2020				6.79†	

Notes: Annual panels begin in 1995; the public-supply panel begins in 2000. The instrumental variable is s_cG_t. Every specification includes county and state-by-year fixed effects; standard errors are clustered by state. Treatment is IHS cumulative annualized county data-center revenue, measured in millions of dollars. Outcomes use IHS except for logged house and electricity prices. At positive revenue values, the electricity-price coefficient is an approximate elasticity. A dagger marks the electricity-price and public-supply specifications, whose inference uses the weak-IV-robust Anderson–Rubin test. Column (6) evaluates each coefficient at the unweighted mean sample change in IHS cumulative revenue among the outcome's estimation counties. Column (7) evaluates the median positive county-by-entry-year revenue increment at the mean positive-revenue base; this is approximately equivalent to an increase of \$10 million at a \$100 million revenue base. This is a mechanical translation of the coefficient, not an estimate of heterogeneity. First-stage statistics can differ despite identical start and end years because of the estimation sample. Sample sizes are 45,733 county-years for each CBP outcome, 44,154 for each IRS outcome, 41,248 for house prices, and 45,079 for electricity prices. The public-supply sample contains 33,117 county-years in 1,577 counties. We denote $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.10$.

hyperscale facilities. Its first stage is therefore weighted toward county revenue growth induced by large, scale-oriented developments. These facilities place high local demands. In constrast, traditional enterprise and many third-party or colocation facilities are more closely tied to established users and urban demand. Consistent with this distinction, Fang and Greenstein (2025) find a stronger urban bias and greater sensitivity to buyer proximity among third-party providers than among major cloud providers, which concentrate in fewer, lower-density locations. Energy and industry evidence also documents the shift toward large hyperscale sites and the importance of scalable power and connectivity for their location (Shehabi et al., 2016; CBRE, 2026).

This distinction offers one plausible interpretation of the large magnitudes of the estimates: our historical instrument is designed to capture the local footprint of the large cloud- and AI-relevant developments that expanded during the sample, rather than the average effect of a smaller enterprise-oriented facility. *These are also the developments likely to become increasingly important as AI raises demand for concentrated compute* (International Energy Agency, 2025). 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.[11] On the common support, its total-employment coefficient is roughly one-tenth of the baseline estimate of; most other point estimates are also smaller.

5 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, $\widetilde{Z}_c^{AN}$, and terrain-route proximity, Z_c^{LC}.

For each endpoint t, we estimate

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

where $\Delta_t Y_c = Y_{ct} - Y_{c,1995}$ and $\Delta_t D_c = D_{ct} - D_{c,1995}$. 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 for economic levels measured in 1986 and 198 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 5 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, it should predict these earlier economic levels. All outcomes are measured between 1985 and 1990. The left panel conditions on historical population and geography, and the right panel 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 Figures 6 and 7 add a second set of placebo checks using changes in the outcomes before 1995. These tests, in orange, document that historical corridor geography did not

[11] Appendix Tables C.2–C.9 report the alternative exposures' early-level and pre-1995 placebo relationships.

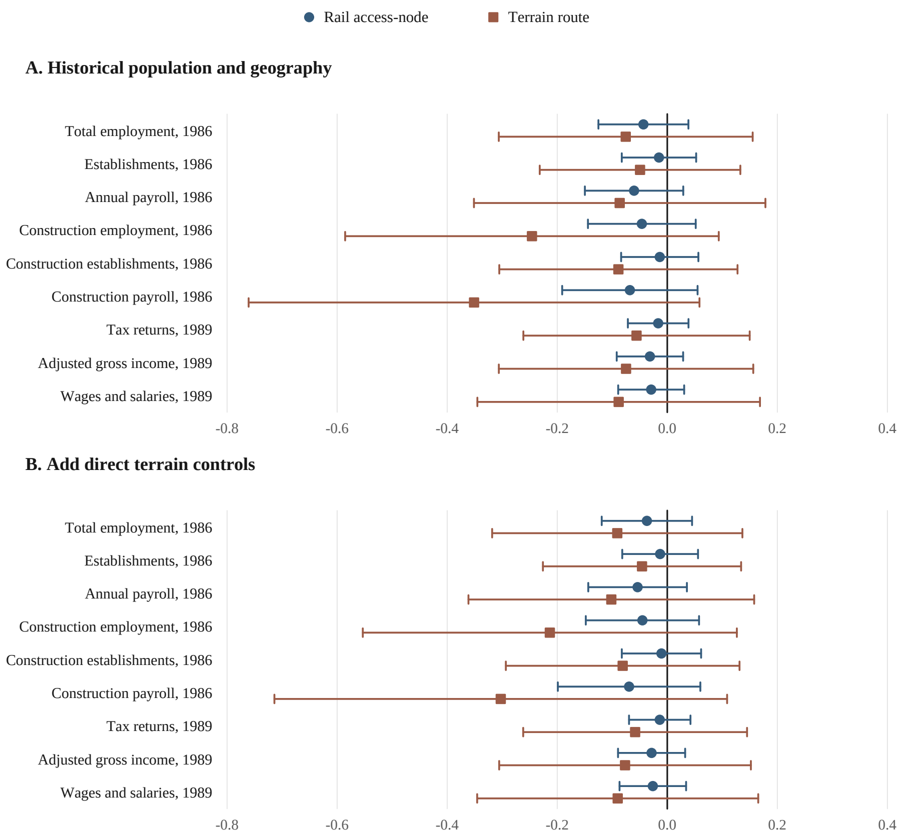

Figure 5: Balance on early county outcomes

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.

generate employment, income, housing, or electricity-price trends before data centers expanded.[12]

5.1 Dynamic long-difference results

The main goal of this section is to document the dynamic responses of each outcome. In each long-difference estimation, the post-period coefficient answers a simple question: by that endpoint, 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 points ask whether the same geography predicted outcome changes before modern data-center expansion. The resulting figures therefore show when local responses emerge.

Table 2 summarizes the estimates at the *latest* endpoint available for each outcome. Figures 6 and 7 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

[12]Because cumulative 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.

intervals (shown by vertical bars) and Anderson–Rubin inference; filled markers indicate AR $p < 0.10$, while open markers indicate AR $p \geq 0.10$.[13]

Three remarks are in order. First, the pre-period points consistently cluster near zero. 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 3. Third, later estimates are consistently positive across the outcomes.

Table 2: Historical-corridor 2SLS estimates at the latest available endpoint

					Implied effect at	
Outcome	End	Estimate	(SE)	p-value	Mean sample change	Median increment at mean positive-revenue base
	(1)	(2)	(3)	(4)	(5)	(6)
Total employment	2023	0.274^{**}	(0.122)	0.031	6.0%	2.6%
Construction employment	2023	0.600^{*}	(0.314)	0.064	13.7%	5.9%
Establishments	2023	0.226^{**}	(0.097)	0.025	5.0%	2.2%
Annual payroll	2023	0.360^{**}	(0.145)	0.018	8.0%	3.5%
Tax returns	2022	0.108	(0.083)	0.200	2.3%	1.0%
Adjusted gross income	2022	0.231^{**}	(0.113)	0.047	5.0%	2.2%
Wages and salaries	2022	0.190^{*}	(0.098)	0.061	4.1%	1.8%
House-price-index growth	2023	$0.442^{***\dagger}$	(0.179)	< 0.001	13.3 pp	4.2 pp
Electricity-price elasticity	2023	$0.048^{\dagger}$	(0.063)	0.606	1.0%	0.5%
Public-supply withdrawals	2020	0.311^{*}	(0.163)	0.065	6.7%	3.0%
Outcome sample	Start	End		First-stage F		
CBP outcomes	1995	2023		11.92		
IRS outcomes	1995	2022		11.87		
House prices	1995	2023		$8.11^{\dagger}$		
Electricity prices	1995	2023		$9.39^{\dagger}$		
Public-supply withdrawals	1995	2020		11.61		

Notes: Each row is a separate 2SLS regression using the rail-coverage-by-access-node interaction and terrain-predicted route proximity jointly. All models use Early + terrain controls, state fixed effects, and state-clustered standard errors. The endpoint is the final year available in the source: 2023 for CBP, FHFA, and EIA outcomes and 2022 and 2020 for IRS outcomes and public supply, respectively. A dagger marks an outcome sample with first-stage $F < 10$. For those rows, the p-value and significance stars use the joint Anderson–Rubin test, which remains valid under weak identification; they do not use the conventional 2SLS t-test. Column (5) evaluates each coefficient at the unweighted mean sample change in IHS cumulative revenue among the outcome's estimation counties. Column (6) evaluates the median positive county-by-entry-year revenue increment at the mean positive-revenue base; this is approximately equivalent to an increase of $10 million at a $100 million revenue base. This is a mechanical translation of the coefficient, not an estimate of heterogeneity. Electricity prices enter in natural logs, so at positive revenue values their coefficient is an approximate elasticity. Rows with $F \geq 10$ use conventional clustered 2SLS inference. We denote $^{***}p < 0.01$, $^{**}p < 0.05$, and $^{*}p < 0.10$.

To grasp the magnitudes of Table 2, consider the 2023 total-employment coefficient of 0.274. This implies that 10 percent more cumulative 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 (5). Column (6) evaluates the median positive revenue increment at the mean positive-revenue base and implies 2.6 percent more employment. Figures 6 and 7 use the same coefficient units. The 2023 total-employment point is 0.259 with Early controls and 0.274 with Early + terrain controls. Evaluated at the same mean revenue change of 23.8

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

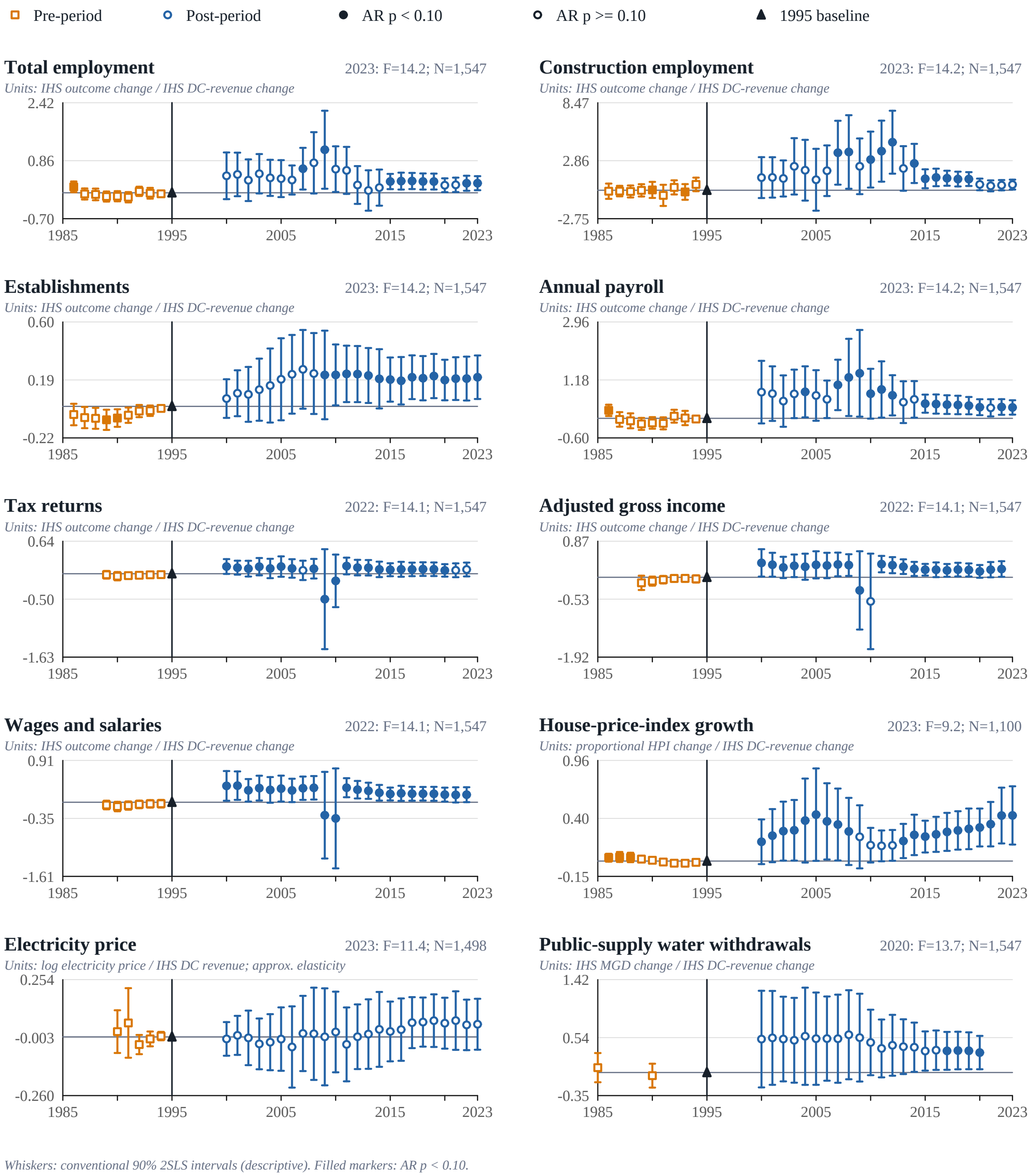

Figure 6: Annual historical-corridor outcome profiles: Early controls

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. Filled markers have Anderson–Rubin $p < 0.10$ and open markers have Anderson–Rubin $p \geq 0.10$. For IHS outcomes the vertical axis has the same approximate elasticity interpretation at positive values; house-price values are proportional-growth effects, and electricity-price values use log prices and therefore have an approximate elasticity interpretation at positive revenue values.

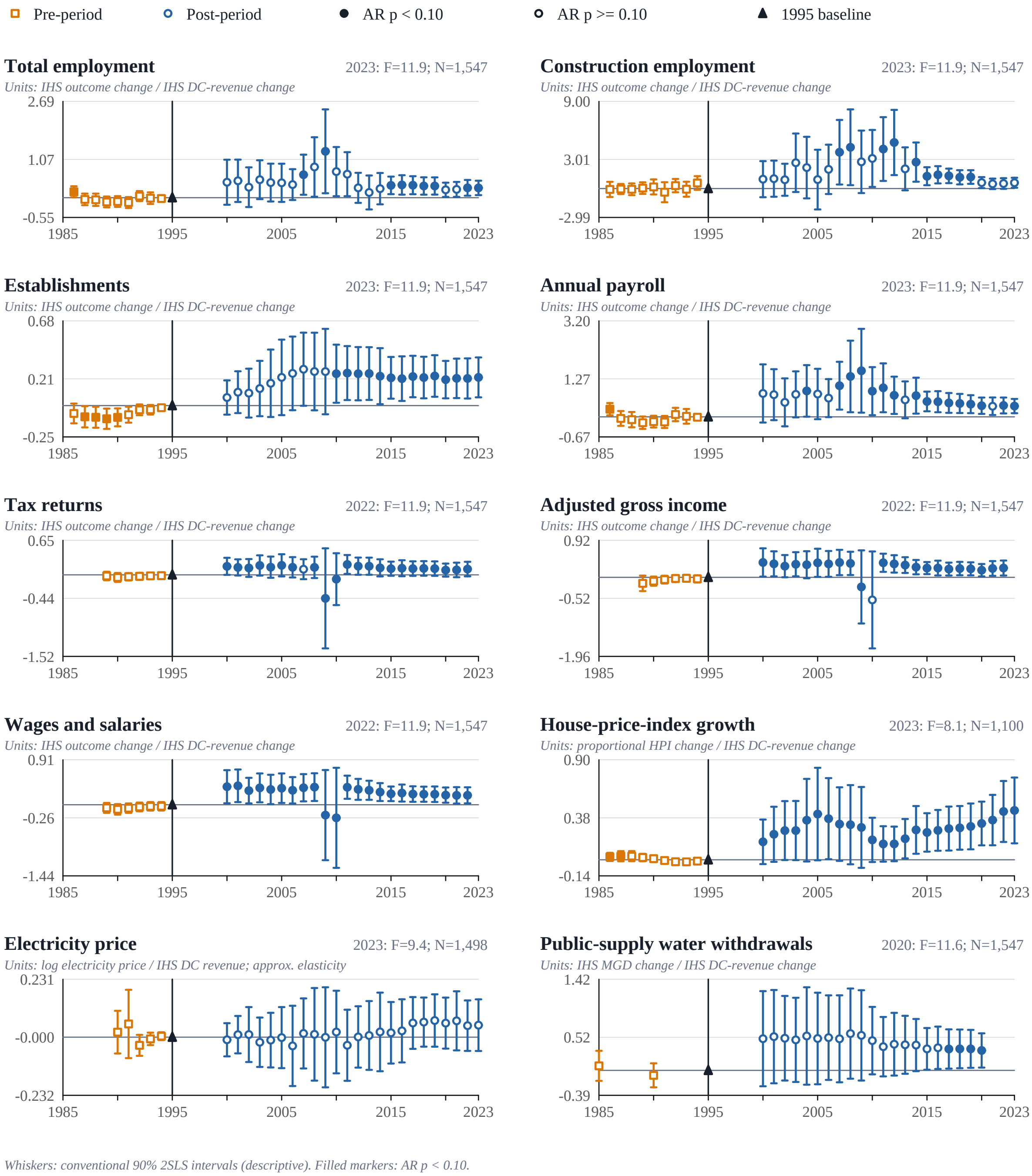

Figure 7: Annual historical-corridor outcome profiles: Early + terrain controls

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. Filled markers have Anderson–Rubin $p < 0.10$ and open markers have Anderson–Rubin $p \geq 0.10$. For IHS outcomes the vertical axis has the same approximate elasticity interpretation at positive values; house-price values are proportional-growth effects, and electricity-price values use log prices 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.

percent, these coefficients imply employment effects of 5.7 and 6.0 percent, respectively.[14] 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. House-price growth is positive and stable across years. The log electricity-price coefficients are also positive in the later years, although they are imprecisely estimated under weak-IV-robust inference.[15]

Public-supply withdrawals Total public-supply withdrawals are positive at later endpoints. For instance, in Figure 7, the 2020 coefficient is 0.311. Thus, a 10 percent increase in cumulative 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.

Property-tax revenue. Additionally, Appendix G explores whether data-center development affects property-tax revenue in hosting counties. Coefficients are generally positive but imprecisely estimated across specifications.[16] Under early controls, estimates become significant at the 10 percent level toward the end of the sample. For instance, at positive values, the 2023 early-controls IV point estimate implies that a 10 percent increase in cumulative reported annualized data-center revenue induced by the historical-corridor instruments is associated with approximately 14 percent greater property-tax-revenue growth between 1995 and 2023.

5.2 Potential Spillovers: The West Des Moines Case Study

The magnitude of our estimates suggests that the direct activity of data centers may not be the entire mechanism. One possibility is that a large data-center investment creates fiscal capacity to finance

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

[15]For 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.

[16]One reason for the limited precision may be that property-tax abatements are common among large-scale facilities (Gargano and Giacoletti, 2025). Based on Bommarito and Bommarito (2026), about 40 percent of U.S. hyperscale projects that were operational, expanding, or under construction had an explicitly documented property-tax abatement. Moreover, higher house prices may expand the local property-tax base even when the data centers themselves receive abatements.

shared infrastructure, lowering the cost of subsequent development and firm entry. West Des Moines provides an unusually clear illustration because the links in this propagation mechanism appear in the city's own fiscal and contractual records.

Microsoft began planning its first data center in West Des Moines in 2008. In 2016, when the city announced Project Osmium—Microsoft's third local campus—it projected an investment of \$1.5–\$2.0 billion and stated that the resulting property-tax revenue would finance streets, sanitary sewer, and water lines. The city expected these investments to accelerate development across more than 5,000 acres (City of West Des Moines, 2016). The infrastructure was therefore not limited to the data-center campus. By FY2025, city records attributed approximately \$285 million in roads, utilities, and fiber infrastructure to tax increment generated by successive data-center projects, opening thousands of acres to additional development (City of West Des Moines, 2025; City of West Des Moines Community and Economic Development Department, 2025).

The data-center investment also initiated broader coordination in the local connectivity market. In 2018, Microsoft approached the city about investing in community connectivity and funded an expansion of Valley Junction Wi-Fi. The project brought together Aureon, which supplied access to its fiber; Ovation Networks, which operated the system; and T-Mobile, which participated in a school-connectivity component (City of West Des Moines, 2018).

The sequence provides a potential rationale for large local effects. A data-center investment expanded the local tax base; the city converted the resulting revenue into infrastructure with uses beyond the campuses; an open-access design lowered the fixed cost of entry; and identifiable broadband providers entered. Building and operating this infrastructure creates demand for construction, installation, maintenance, network operations, and customer service, while improved connectivity may reduce costs for other establishments.

5.3 A Spatial-Equilibrium Interpretation

In this section we sketch the elements of a simple model that can rationalize the qualitative patterns of our estimates. In the model we view a data center as a local export industry. It produces computing services for customers that may be located anywhere, but it relies on local workers, electricity, water, and land. Worker chose locations where to live and work, subject to location specific preference shocks. Besides the idiosyncratic location taste, workers consume local housing services and receive wages. Computing services, as well as other tradable goods, use labor and other local services as inputs. Housing, electricity, and water have upward-sloping supply curves at the local level, reflecting fixed factors such as land and infrastructure. A model like this is the modern incarnation of the canonical spatial-equilibrium model in the spirit of Roback (1982). We interpretate locations in the model as counties in the data.

We assume location are heterogeneous on the suitability to install a data center. We sketch here two possible comparative static exercises that in such a model rationalizes the entry of a data center in a location. The first one is a technological development in either the demand of data center services –say an increase in their price, which makes some locations suitable for the installation of the data center. The second will be a fiscal subsidy for the installation of a data center. In both cases, when a data center enters in a location it produces an increase in the demand for local inputs, which puts upward pressure on wages as well as other inputs. Higher wages attract workers from other locations, and higher population levels and incomes increases demand for housing and local services. Greater demand for electricity and water puts upward pressure on prices. The effects on employment, housing, and infrastructure are therefore different parts of the same local adjustment (Roback, 1982; Moretti, 2010b). The comparative statics of this type of model match the qualitative patterns of our estimates across counties.

Because data centers are concentrated in a relatively small number of counties, as a fist passage, these local effects can be understood relative to an outside economy that determines workers' alternative opportunities, which are themselves not affected by the installation of data centers.

We discuss the different implications of the two reasons for the enter of a data center in the model. Both have the same qualitative implications for the variables that we end up measuring in our data, but in the model the entry of a data center would have different effects on welfare. In the benchmark case of these models, which features no externalities –beyond pecuniary ones– the technological driven explanation will be welfare improving – or more precisely it will attract workers to the locations until it dissipate the marginal welfare effects, and produce gains on the owner of fixed factors in such locations. In the second case, i.e. in the subsidy driven explanation of the data center entry, the effect on welfare in negative – the gains do not compensate the losses from tax revenues. Adding externalities –beyond pecuniary ones– to these baseline cases can change both results, depending on the balance of agglomeration and congestions effects.

The West Des Moines case suggests that a mixed between the two explanation may be at play. It also suggest a more nuanced view of the local subsides to infrastructure. If the data-center tax base allows the local government to finance roads, water and sewer systems, or fiber infrastructure, the investment may increase the profitability of other firms, further expanding the amount of land that can be developed, or lower the fixed cost of entry. These effects can magnify the effect on activity relative to more standard cases (Kline and Moretti, 2014; Suárez Serrato and Zidar, 2016). Thus, this spillover raises local demand beyond the direct effect, so that the extra public investment can further increase employment, establishment entry, payroll, and income. Their net welfare effect, absent of externalities, will be also negative, i.e. will not compensate the loss on revenue.

In the simple model, wages rise and workers move in from other locations in response to the direct and indirect effect on labor demand, house prices rise, new establishments enter, and employment and the local wage bill increase. Greater demand for electricity and water also raises their use and can put upward pressure on prices when local supply is constrained. At the same time, infrastructure investments that expand housing or utility capacity may amplify quantity responses while moderating some price increases. These patters are consistent with the qualitative patters of our estimates, which show increases in employment, establishments, payroll and income, house prices, and public-supply water withdrawals, as well as the increase in electricity prices at longer horizons. Our reduced-form estimates capture the combined local-equilibrium response and do not separately identify the direct demand and fiscal-infrastructure channels. We also suspect that the employment and activity effect that we estimate are too high relative to most existing geography models. We leave the analysis of these consideration for future research.

6 Conclusion

This paper studies how data-center development affects local employment, business activity, income, house prices, electricity prices, and public-supply water withdrawals. The analysis combines a facility-level panel of U.S. data centers with county outcomes and uses historical corridor geography to address endogenous site selection.

The identification strategy starts from the idea that data-center suitability depends on two margins of high-capacity connectivity. Proximity to a feasible long-distance route provides a baseline advantage, while a usable access point raises the value of that route by allowing traffic to enter, leave, or branch from the network. We recover the baseline route margin using proximity to terrain-predicted transcontinental routes between fixed 1855 terminal pairs. We recover the additional access margin using the interaction of concentrated 1870 rail coverage with internal endpoints and junctions. A design that pools outcomes between 2000 and 2023 combines these measures into a county exposure share and interacts it with leave-U.S.-out growth in scale-oriented facility openings. A dynamic long-difference design uses the same two margins to recover dynamic effects while focusing on longer-run changes.

The analysis therefore introduces a new county-level historical rail-access instrument that combines corridor concentration with internal access-node orientation. To the best of our knowledge, this is the first application of historical rail or network topology to predict data-center development.

The variation isolated by our instrument is weighted toward large, expandable facilities rather than traditional enterprise data centers. These are precisely the sites most likely to expand with AI, because training and inference require concentrated computing capacity, high-capacity connectivity, reliable power, and room to add capacity. Understanding their local effects is important for decisions over permitting, land use, grid investment, and tax incentives, all of which can commit local governments and utilities to long-lived policy and infrastructure choices.

The long-difference estimates show positive effects on total employment, construction employment, establishments, payroll, tax returns, adjusted gross income, and wages. For scale, at positive revenue values, a 10 percent increase in the cumulative reported annualized revenue of facilities built in a county implies approximately 2.7 percent more total employment, 6.0 percent more construction employment, 2.3 percent more establishments, and 3.6 percent more payroll between 1995 and 2023. The corresponding 1995–2023 estimates imply 0.5 percent higher electricity prices and an additional 4.4 percentage points of house-price growth. At the latest available endpoints, the estimates imply 1.1 percent more tax returns, 2.3 percent more adjusted gross income, and 1.9 percent more wages through 2022, and 3.1 percent more public-supply water withdrawals through 2020. The dynamic estimates show that employment and establishment gains persist, construction effects are strongest at shorter horizons, house-price increases are sustained, and positive, though less precise, electricity-price estimates emerge later. Public-supply water withdrawals also rise, consistent with an additional local infrastructure demand created by data-center development.

Evaluations of tax incentives, permitting, and infrastructure cost recovery should therefore weigh gains in employment, business activity, value of the housing stock, and income against pressure on power and water markets, while also considering who bears the cost of the infrastructure required by these facilities. As explained in the last section, we will require more detailed empirical analysis of the determinants of the entry decision to attempt a serious welfare evaluation. Moreover, our analysis has been limited to the county level effects. This is a serious limitation since it is likely that the effect vary differently at the very short distances relative to larger distances but still in the same labor and housing markets. We leave these important questions for future research. We believe that the issues analyzed in this paper, as well as these further questions will become increasingly important as AI raises demand for scale-oriented computing infrastructure.

References

Akerman, A., Gaarder, I., Mogstad, M., 2015. The skill complementarity of broadband internet. Quarterly Journal of Economics 130, 1781–1824.

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.

Bartik, T.J., 1991. Who Benefits from State and Local Economic Development Policies? W.E. Upjohn Institute for Employment Research, Kalamazoo, MI.

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.

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

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

CBRE, 2025. North america data center trends h2 2025. URL: https://www.cbre.com/insights/books/north-america-data-center-trends-h2-2025. cBRE Research.

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

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

City of West Des Moines, 2018. City of west des moines and microsoft partner to expand wi-fi access in valley junction. City News and Updates. City of West Des Moines, Iowa.

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

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

Donaldson, D., Hornbeck, R., 2016. Railroads and american economic growth: A market access approach. Quarterly Journal of Economics 131, 799–858. doi:10.1093/qje/qjw002.

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

Faber, B., 2014. Trade integration, market size, and industrialization: Evidence from china's national trunk highway system. Review of Economic Studies 81, 1046–1070. doi:10.1093/restud/rdu010.

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

Fenske, J., Kala, N., Wei, J., 2023. Railways and cities in india. Journal of Development Economics 161, 103038. doi:10.1016/j.jdeveco.2022.103038.

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

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

Gargano, A., Giacoletti, M., 2025. Subsidizing the Cloud: U.S. State Incentives to Data Centers. SSRN Working Paper. SSRN. URL: https://ssrn.com/abstract=5881105, doi: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.

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.

Hogendorn, C., 2011. Excessive(?) entry of national telecom networks, 1990–2001. Telecommunications Policy 35, 920–932. doi: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, 2025. Energy and AI. IEA report. Analysis of electricity demand from data centres and artificial intelligence.

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.

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., 2010a. Local multipliers. American Economic Review: Papers and Proceedings 100, 373–377.

Moretti, E., 2010b. Local multipliers. American Economic Review 100, 373–377. doi:10.1257/aer.100.2.373.

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.

Online Appendix for

Data Centers and Local Economies in the Age of AI: A Shift–Share Approach

Fernando Alvarez
University of Chicago and NBER

David Argente
Yale University and NBER

Joyce Chow
Yale University and NBER

Diana Van Patten
Yale University and NBER

A Construction and estimating equations

This appendix records the exact transformations summarized in words in the main text.

A.1 Treatment

Let r_i denote the annualized revenue currently reported for facility i, b_i its reported build year, and $c(i)$ its county. We define the cumulative reported scale of facilities built in county c by year t as

$$Q_{ct} = \sum_i \mathbf{1}\{c(i) = c,\ 1995 \le b_i \le t\} r_i, \qquad D_{ct} = \operatorname{asinh}(Q_{ct}), \qquad \Delta_t D_c = D_{ct} - D_{c,1995}. \tag{6}$$

Equivalently, if

$$R_{cs} = \sum_i \mathbf{1}\{c(i) = c,\ b_i = s\} r_i,$$

then $Q_{ct} = \sum_{s=1995}^{t} R_{cs}$. Each facility's currently reported annualized revenue enters the cumulative county scale beginning in its reported build year. The pooled annual panel uses D_{ct}; the endpoint-specific long differences use $\Delta_t D_c$.

A.2 Rail access-node and terrain-route components

A fixed 5-km grid is placed inside each historical county. Let G_c be the number of grid points and $S_c(h)$ the share within h kilometers of the 1870 rail network. We form smoothed log odds and the near-versus-broad coverage measure

$$L_c(h) = \log\left[\frac{G_c S_c(h) + 0.5}{G_c\{1 - S_c(h)\} + 0.5}\right], \qquad C_c = L_c(10) - L_c(50). \tag{7}$$

The half-cell adjustment keeps the log odds finite when all or none of a county's grid points fall within a band. The measure is higher when broad rail exposure is concentrated close to the line.

Let N_c count endpoints and junctions inside the county, excluding ordinary degree-two points where a line simply continues, and let B_c count distinct places where a rail line crosses the county boundary. Access-node geometry is

$$A_c = \log(1 + N_c) - \log(1 + B_c). \tag{8}$$

We standardize the two lower-order components as

$$\widetilde{C}_c = \frac{C_c - \bar{C}}{s_C}, \qquad \widetilde{A}_c = \frac{A_c - \bar{A}}{s_A}, \tag{9}$$

where $\bar{C}$ and $\bar{A}$ are their means and s_C and s_A are their standard deviations in the instrument-construction sample. Thus, throughout this appendix, a tilde explicitly denotes a standardized variable. We first define the rail access-node interaction before its own standardization as

$$Z_c^{AN} = \widetilde{C}_c \widetilde{A}_c. \tag{10}$$

Its standardized version, which is used as the rail access-node instrument, is

$$\widetilde{Z}_c^{AN} = \frac{Z_c^{AN} - \overline{Z^{AN}}}{s_{Z^{AN}}}. \tag{11}$$

Here, $\overline{Z^{AN}}$ and $s_{Z^{AN}}$ are the mean and standard deviation of Z_c^{AN} in the instrument-construction sample. The tilde on $\widetilde{Z}_c^{AN}$ therefore states explicitly that the interaction itself is standardized.

The long-difference and balance regressions include $\widetilde{C}_c$ and $\widetilde{A}_c$ separately; county fixed effects absorb their time-invariant levels in the pooled panel. Equivalently, $A_c = -[\log(1 + B_c) - \log(1 + N_c)]$. Writing the ratio in the opposite direction would reverse the displayed rail coefficient but leave the instrument and control spans, fitted treatment, first-stage statistics, and 2SLS estimates unchanged.

The terrain construction uses six terminal pairs: St. Paul–Puget Sound, Council Bluffs–Sacramento, St. Joseph–San Francisco, Fort Smith–San Pedro, Vicksburg–San Diego, and San Antonio–San Diego. Moving across an edge e of the terrain grid has cost

$$K_e = d_e \left(1 + 400 g_e^2\right), \tag{12}$$

where d_e is horizontal distance and g_e is absolute grade calculated from the ETOPO bedrock-elevation surface. An A-star routine selects the minimum-cost route for each pair. If d_c^{LC} is the shortest straight-line distance from the county centroid to one of the six predicted paths, terrain-route proximity is

$$Z_c^{LC} = \operatorname{std}\left\{-\log(1 + d_c^{LC})\right\}. \tag{13}$$

A.3 The scale-oriented exposure share and scale-growth shift

With the measures written in the economic directions defined above, we form the fixed equal-weight index

$$H_c = \frac{\widetilde{Z}_c^{AN} + Z_c^{LC}}{\sqrt{2}}. \tag{14}$$

The scale-oriented exposure share is an affine rescaling of this index,

$$s_c = \frac{H_c - \min_j H_j}{\max_j H_j - \min_j H_j}. \tag{15}$$

Both standardized lower-order components can be positive or negative. Because the two components are centered before they are interacted, Z_c^{AN} can be positive when both components are above their respective means or when both are below them. This is a standard property of a centered interaction. Algebraically,

$$\widetilde{C}_c \widetilde{A}_c = \frac{(C_c - \bar{C})(A_c - \bar{A})}{s_C s_A} = \frac{C_c A_c - \bar{A} C_c - \bar{C} A_c + \bar{C}\bar{A}}{s_C s_A}.$$

When the lower-order terms are included, using the centered interaction is equivalent, up to an intercept and linear terms in C_c and A_c, to using the uncentered interaction $C_c A_c$. Thus, in the long-difference and balance specifications, high–high and low–low counties can have the same interaction component, but they need not have the same fitted value because their linear components differ and enter separately as controls. Standardizing the interaction to form $\widetilde{Z}_c^{AN}$ changes only its location and scale.

The interaction coefficient captures a cross-partial relationship: whether the predictive relationship between concentrated historical rail coverage and subsequent data-center development varies with access-node orientation. It should not be interpreted as a monotone index in which a larger value necessarily means that both underlying components are high. In the pooled shift-share specification, $\widetilde{Z}_c^{AN}$ is a prespecified component of the county exposure and retains this cross-partial interpretation.

The terrain component separately captures proximity to a low-cost long-distance route. We place the upper end of their equal-weight ranking on the scale-oriented side because large wholesale, powered-shell, and hyperscale campuses make intensive use of backbone capacity, route diversity, and infrastructure that can expand with their power, cooling, and computing needs.

The foreign shift uses 1,457 scale-oriented facilities in 54 countries outside the United States, classified by reported build year within the 1995–2023 cohort. Scale-oriented facilities include Wholesale, Powered Shell, and Hyperscale sites. We include facilities currently classified as operational or not operational, so the measure captures cumulative openings rather than only surviving sites. Planned, under-construction, land-held, and cryptocurrency facilities are excluded.
Define

$$N_t = \sum_{\tau=1995}^{t} \#\{\text{foreign scale-oriented facilities built in } \tau\}, \qquad G_t = \operatorname{asinh}(N_t) - \operatorname{asinh}(N_{1995}). \quad (16)$$

The excluded variable used in the pooled annual design is defined directly as

$$Z_{ct} = s_c G_t. \quad (17)$$

The county weight determines where foreign expansion should matter, and G_t determines when it should matter. The instrument therefore rises most near the upper end of the historical ranking as scale-oriented facilities accumulate abroad. Because the shift is constructed entirely from facilities outside the United States, no U.S. facility or county outcome enters it.

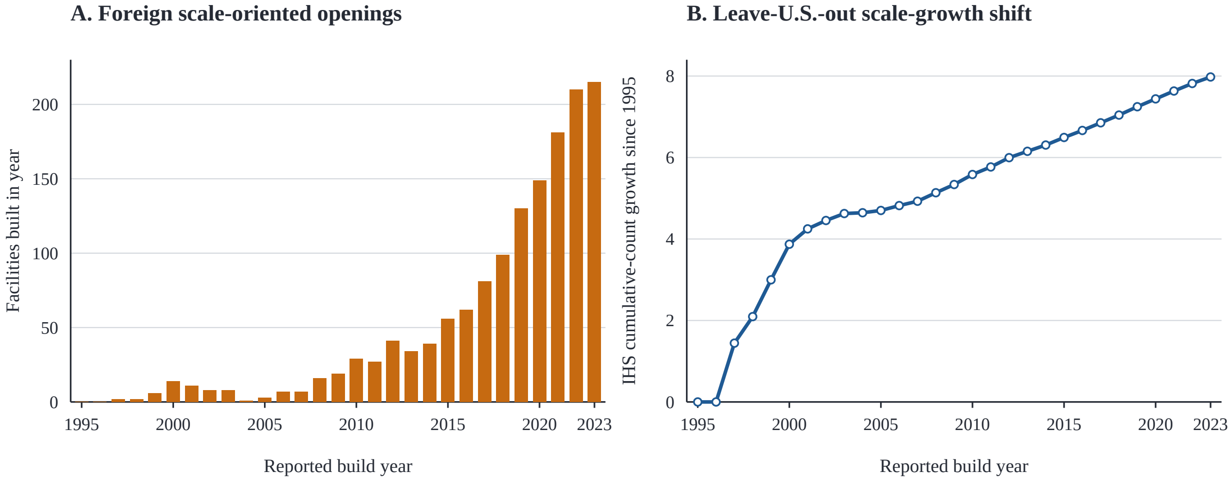

Figure A.1: Foreign growth in scale-oriented facilities

Notes: Panel A plots annual leave-U.S.-out openings of wholesale, powered-shell, and hyperscale facilities. Panel B plots the cumulative-IHS scale-growth shift G_t used in equation (2). Both series are zero in 1995 because the first qualifying foreign cohorts are built in 1997. Annual openings rise from 34 in 2013 to 215 in 2023. The shift equals 6.153 in 2013 and 7.977 in 2023; these are IHS cumulative-count growth units, not percentages. No U.S. facility or county outcome enters either series. Source: 451 Research Datacenter Knowledge Base.

A.4 Estimating equations and controls

The preferred pooled annual design uses equations (3)–(4) with $\mathbf{Z}_{ct} = Z_{ct}$. The fixed effects, outcome transformations, clustering, and interpretation of β are described in Section 4.

For endpoint t, the dynamic joint first stage is

$$\Delta_t D_c = \pi_{1t}\widetilde{Z}_c^{AN} + \pi_{2t}Z_c^{LC} + \rho_{1t}\widetilde{C}_c + \rho_{2t}\widetilde{A}_c + X_c'\gamma_t + \alpha_s + v_{ct}, \tag{18}$$

and the second stage is the compact specification shown in the main text. "Early controls" comprise log 1870 population, log land area, water proximity, terrain ruggedness, historical temperature, the separate rail-coverage and access-node-geometry measures, 1986 IHS total employment, establishments, and payroll, 1989 IHS tax returns, adjusted gross income, and wages, and state fixed effects. "Early + terrain" adds mean log nonnegative elevation, within-county elevation dispersion, and mean slope. Neither set includes 1980 population, the 1980 urban-college share, or 1995 economic outcomes.

B First-stage details

This appendix reports the joint first-stage coefficients and their evolution across endpoints. The table provides selected benchmark endpoints supporting the main-text first-stage discussion, and the figure displays the same pattern graphically.

Table B.1: Joint first stage for cumulative data-center revenue

Controls	Endpoint	Rail access-node	Terrain route	Joint F	Joint p	Partial R^2	N
Early	2000	0.028 (0.018) $p = 0.121$	0.029 (0.035) $p = 0.420$	1.85	0.172	0.003	1,547
	2005	0.029 (0.018) $p = 0.116$	0.038 (0.039) $p = 0.329$	2.14	0.131	0.003	1,547
	2010	0.040 (0.023) $p = 0.089$	0.067 (0.036) $p = 0.069$	4.73	0.015	0.005	1,547
	2015	0.055 (0.022) $p = 0.017$	0.117 (0.035) $p = 0.002$	10.44	0.000	0.008	1,547
	2020	0.070 (0.025) $p = 0.008$	0.136 (0.040) $p = 0.001$	13.74	0.000	0.011	1,547
Early + terrain	2000	0.029 (0.018) $p = 0.120$	0.024 (0.035) $p = 0.501$	1.79	0.180	0.003	1,547
	2005	0.030 (0.018) $p = 0.109$	0.029 (0.038) $p = 0.446$	2.05	0.143	0.003	1,547
	2010	0.041 (0.023) $p = 0.083$	0.055 (0.035) $p = 0.128$	4.33	0.020	0.005	1,547
	2015	0.058 (0.023) $p = 0.014$	0.093 (0.039) $p = 0.021$	8.05	0.001	0.007	1,547
	2020	0.072 (0.025) $p = 0.006$	0.115 (0.042) $p = 0.010$	11.61	0.000	0.010	1,547

Notes: Rail access-node is standardized 1870 rail coverage ELD(10,50) interacted with standardized access-node geometry, defined as internal endpoints and junctions relative to boundary crossings; both lower-order components enter as controls. Terrain route is standardized negative log distance to the 1855 terrain-predicted least-cost routes. Joint F is the state-clustered test that both historical-corridor coefficients are zero. Bold denotes $F \geq 10$. Appendix Tables F.1–F.4 report all 48 joint balance regressions; none rejects at 5 percent. To read the first coefficient, a one-standard-deviation increase in the rail access-node measure predicts a 0.028-IHS-point larger 1995–2000 increase in cumulative data-center revenue—approximately 2.8 percent at positive revenue values—holding the terrain instrument, lower-order components, and controls fixed.

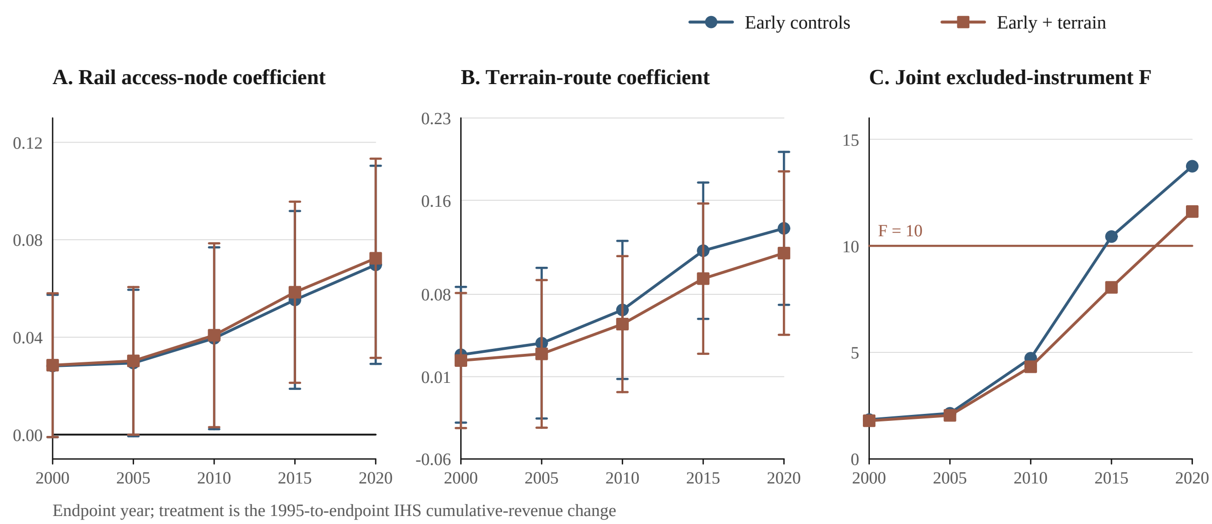

Figure B.1: The joint first stage strengthens with the cumulative horizon

Notes: Points are first-stage coefficients and interval bars are 90 percent confidence intervals using state-clustered standard errors. Panel C reports the joint first-stage statistic in the common sample. To read the first point in Panel A, a one-standard-deviation increase in the rail access-node measure predicts a 0.028-IHS-point larger 1995–2000 increase in cumulative data-center revenue—approximately 2.8 percent at positive revenue values. The associated joint first-stage statistic is 1.85, so this early endpoint is weakly identified.

C Alternative shift-share benchmark

The alternative approach follows a shift-share logic (Bartik, 1991). Let F_c denote negative log distance to a pre-existing long-haul fiber node mapped by Durairajan et al. (2015), and let U_c denote the county's 1980 share of the national urban-college population. Let G_t^{ROW} denote cumulative IHS data-center revenue growth since 1995 in the rest of the world outside the United States and China. We construct two excluded variables,

$$Z_{ct}^F = F_c G_t^{ROW}, \qquad Z_{ct}^U = U_c G_t^{ROW}. \tag{19}$$

Both excluded variables use the same foreign growth series. Rest-of-world growth supplies the common timing of an external industry expansion, while the two county measures allocate that shift differently across space. The construction's intuition is straightforward: proximity to long-haul fiber captures access to network capacity and redundancy, while historical human capital is plausibly associated with early demand for digital infrastructure.

Table C.1 enters the two excluded variables jointly in the pooled annual specification described in Section 4. The coefficient β is one pooled summary of the relationship between accumulated data-center development and each outcome. The alternative exposures are less plausibly exogenous because modern fiber-node proximity and historical human capital may affect county outcomes through channels other than data-center development. We therefore treat this specification as a descriptive benchmark rather than as the preferred causal design.

The pre-expansion evidence confirms the exclusion concern. Tables C.2–C.5 reproduce the preferred design's complete 12-outcome early-level battery for both historical-population cohorts and both control sets. The alternative exposures jointly predict 11 of the 12 early outcomes at the 5 percent level in every panel, for 44 rejections among 48 tests.

Tables C.6–C.9 reproduce all of the pre-1995 outcome changes used as placebos for the preferred design, holding the endogenous treatment fixed at its 1995–2020 change. Under Anderson–Rubin inference, 16 of 70 placebo tests reject at 5 percent with Early controls and 21 of 70 reject with Early + terrain controls. These pre-expansion relationships reinforce our treatment of the alternative specification as a descriptive benchmark.

Table C.1: Alternative shift-share benchmark: pooled annual estimates

Outcome	End	Estimate	(SE)	First-stage F	Conventional p	AR p
Panel A: Full fiber/college-share support						
Total employment	2023	0.036***	(0.011)	40.95	0.0023	< 0.001
Construction employment	2023	0.836***	(0.089)	40.95	< 0.001	< 0.001
Establishments	2023	0.076***	(0.008)	40.95	< 0.001	< 0.001
Annual payroll	2023	0.019	(0.012)	40.95	0.134	< 0.001
Tax returns	2022	0.059***	(0.013)	40.96	< 0.001	< 0.001
Adjusted gross income	2022	0.079***	(0.012)	40.96	< 0.001	< 0.001
Wages and salaries	2022	0.068***	(0.011)	40.96	< 0.001	< 0.001
House-price index	2023	0.052***	(0.018)	34.24	0.0048	< 0.001
Electricity-price elasticity	2023	0.003	(0.017)	29.69	0.881	0.712
Public-supply water withdrawals	2020	0.064***	(0.015)	22.53	< 0.001	< 0.001
Panel B: Preferred-design common support						
Total employment	2023	0.042***	(0.010)	34.72	< 0.001	< 0.001
Construction employment	2023	0.756***	(0.071)	34.72	< 0.001	< 0.001
Establishments	2023	0.056***	(0.009)	34.72	< 0.001	< 0.001
Annual payroll	2023	0.031***	(0.011)	34.72	0.0098	< 0.001
Tax returns	2022	0.053***	(0.011)	34.90	< 0.001	< 0.001
Adjusted gross income	2022	0.052***	(0.011)	34.90	< 0.001	< 0.001
Wages and salaries	2022	0.049***	(0.010)	34.90	< 0.001	< 0.001
House-price index	2023	0.017	(0.010)	34.30	0.101	0.205
Electricity-price elasticity	2023	-0.014	(0.012)	27.26	0.241	0.467
Public-supply water withdrawals	2020	0.065***	(0.015)	12.21	< 0.001	< 0.001

Notes: The two excluded variables are $F_c G_t^{ROW}$ and $U_c G_t^{ROW}$, entered jointly and defined in equation (19). Both use the same cumulative-IHS data-center revenue growth in the rest of the world outside the United States and China. Every model follows the pooled annual specification in Section 4, with county and state-by-year fixed effects and state-clustered standard errors. Panel A uses all counties with both exposures; Panel B restricts to the preferred historical-corridor support. Outcomes use IHS except for logged house prices and logged electricity prices. The reported first-stage statistic is the joint state-clustered Wald F. Stars use conventional clustered inference because every displayed first stage exceeds 10; weak-IV-robust Anderson–Rubin p-values are also reported. $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$.

Reading the first coefficient: at positive values where IHS approximates a log, the Panel A estimate of 0.036 implies that 10 percent more cumulative data-center revenue is associated with about 0.4 percent more total employment. On the common support, the corresponding estimate is 0.042, or about 0.4 percent more total employment. These are approximate percentage interpretations because IHS also admits zeros.

Table C.2: Alternative-exposure early-level balance: 1860 cohort, historical population and geography

Early outcome	Fiber b	Fiber SE	Fiber p	College b	College SE	College p	Joint F	Joint p	Partial R^2	N
IHS adjusted gross income, 1989	0.391	0.070	< 0.001	0.497	0.137	< 0.001	43.85	< 0.001	0.281	1,270
IHS construction employment, 1986	0.492	0.091	< 0.001	0.595	0.176	0.002	40.59	< 0.001	0.130	1,270
IHS construction establishments, 1986	0.357	0.072	< 0.001	0.458	0.133	0.002	39.28	< 0.001	0.207	1,270
IHS construction payroll, 1986	0.583	0.109	< 0.001	0.648	0.197	0.002	37.42	< 0.001	0.097	1,270
IHS tax returns, 1989	0.320	0.061	< 0.001	0.460	0.126	< 0.001	42.04	< 0.001	0.284	1,270
IHS total annual payroll, 1986	0.308	0.089	0.002	0.592	0.173	0.002	22.14	< 0.001	0.197	1,270
IHS total employment, 1986	0.281	0.076	< 0.001	0.551	0.160	0.002	25.66	< 0.001	0.209	1,270
IHS total establishments, 1986	0.234	0.063	< 0.001	0.487	0.139	0.001	30.15	< 0.001	0.246	1,270
IHS wages, 1989	0.420	0.073	< 0.001	0.497	0.136	< 0.001	44.76	< 0.001	0.279	1,270
Log house-price index, 1986	0.081	0.023	0.001	0.057	0.019	0.006	13.72	< 0.001	0.112	663
Logged electricity price, 1990	0.010	0.013	0.447	0.007	0.004	0.058	2.78	0.076	0.005	1,242
IHS public-supply water withdrawals, 1985	0.209	0.066	0.003	0.494	0.129	< 0.001	25.69	< 0.001	0.196	1,270

Notes: Each row is a separate early-level regression. The two alternative exposures—negative log distance to a long-haul fiber node and the 1980 urban-college share—enter jointly after each is standardized in the full exposure file. Controls, cohort restrictions, outcome transformations, state fixed effects, and state-clustered inference match the corresponding preferred-instrument balance panel. No early economic outcome is included as a control in its own balance regression. In this panel, 11 of 12 joint tests reject at 5 percent.

Table C.3: Alternative-exposure early-level balance: 1860 cohort, add direct terrain controls

Early outcome	Fiber b	Fiber SE	Fiber p	College b	College SE	College p	Joint F	Joint p	Partial R^2	N
IHS adjusted gross income, 1989	0.393	0.071	< 0.001	0.497	0.137	< 0.001	47.55	< 0.001	0.279	1,270
IHS construction employment, 1986	0.500	0.093	< 0.001	0.594	0.176	0.002	42.50	< 0.001	0.130	1,270
IHS construction establishments, 1986	0.354	0.074	< 0.001	0.458	0.133	0.002	39.59	< 0.001	0.206	1,270
IHS construction payroll, 1986	0.605	0.113	< 0.001	0.647	0.198	0.002	39.08	< 0.001	0.098	1,270
IHS tax returns, 1989	0.321	0.062	< 0.001	0.460	0.126	< 0.001	45.80	< 0.001	0.283	1,270
IHS total annual payroll, 1986	0.288	0.091	0.003	0.594	0.173	0.002	22.19	< 0.001	0.194	1,270
IHS total employment, 1986	0.261	0.078	0.002	0.552	0.160	0.001	26.06	< 0.001	0.206	1,270
IHS total establishments, 1986	0.231	0.064	0.001	0.487	0.139	0.001	30.20	< 0.001	0.245	1,270
IHS wages, 1989	0.422	0.074	< 0.001	0.497	0.137	< 0.001	47.74	< 0.001	0.278	1,270
Log house-price index, 1986	0.073	0.022	0.002	0.057	0.019	0.005	13.80	< 0.001	0.104	663
Logged electricity price, 1990	0.014	0.011	0.219	0.007	0.004	0.082	2.77	0.077	0.008	1,242
IHS public-supply water withdrawals, 1985	0.194	0.070	0.009	0.494	0.128	< 0.001	24.92	< 0.001	0.194	1,270

Notes: Each row is a separate early-level regression. The two alternative exposures—negative log distance to a long-haul fiber node and the 1980 urban-college share—enter jointly after each is standardized in the full exposure file. Controls, cohort restrictions, outcome transformations, state fixed effects, and state-clustered inference match the corresponding preferred-instrument balance panel. No early economic outcome is included as a control in its own balance regression. In this panel, 11 of 12 joint tests reject at 5 percent.

Table C.4: Alternative-exposure early-level balance: 1870 cohort, historical population and geography

Early outcome	Fiber b	Fiber SE	Fiber p	College b	College SE	College p	Joint F	Joint p	Partial R^2	N
IHS adjusted gross income, 1989	0.398	0.064	< 0.001	0.504	0.139	< 0.001	49.22	< 0.001	0.261	1,547
IHS construction employment, 1986	0.482	0.092	< 0.001	0.609	0.179	0.002	37.20	< 0.001	0.113	1,547
IHS construction establishments, 1986	0.361	0.068	< 0.001	0.466	0.135	0.001	40.28	< 0.001	0.189	1,547
IHS construction payroll, 1986	0.570	0.113	< 0.001	0.664	0.200	0.002	33.30	< 0.001	0.083	1,547
IHS tax returns, 1989	0.328	0.056	< 0.001	0.467	0.127	< 0.001	46.53	< 0.001	0.263	1,547
IHS total annual payroll, 1986	0.324	0.080	< 0.001	0.600	0.175	0.001	25.47	< 0.001	0.174	1,547
IHS total employment, 1986	0.288	0.069	< 0.001	0.557	0.162	0.001	28.99	< 0.001	0.186	1,547
IHS total establishments, 1986	0.233	0.059	< 0.001	0.495	0.142	0.001	31.03	< 0.001	0.224	1,547
IHS wages, 1989	0.436	0.068	< 0.001	0.504	0.138	< 0.001	48.65	< 0.001	0.262	1,547
Log house-price index, 1986	0.079	0.020	< 0.001	0.058	0.020	0.006	16.72	< 0.001	0.109	760
Logged electricity price, 1990	0.012	0.011	0.289	0.007	0.003	0.052	2.62	0.086	0.006	1,517
IHS public-supply water withdrawals, 1985	0.221	0.060	< 0.001	0.502	0.129	< 0.001	29.06	< 0.001	0.186	1,547

Notes: Each row is a separate early-level regression. The two alternative exposures—negative log distance to a long-haul fiber node and the 1980 urban-college share—enter jointly after each is standardized in the full exposure file. Controls, cohort restrictions, outcome transformations, state fixed effects, and state-clustered inference match the corresponding preferred-instrument balance panel. No early economic outcome is included as a control in its own balance regression. In this panel, 11 of 12 joint tests reject at 5 percent.

Table C.5: Alternative-exposure early-level balance: 1870 cohort, add direct terrain controls

Early outcome	Fiber b	Fiber SE	Fiber p	College b	College SE	College p	Joint F	Joint p	Partial R^2	N
IHS adjusted gross income, 1989	0.400	0.065	< 0.001	0.499	0.138	< 0.001	52.33	< 0.001	0.256	1,547
IHS construction employment, 1986	0.490	0.096	< 0.001	0.598	0.176	0.002	37.87	< 0.001	0.111	1,547
IHS construction establishments, 1986	0.359	0.070	< 0.001	0.460	0.133	0.001	40.92	< 0.001	0.185	1,547
IHS construction payroll, 1986	0.591	0.119	< 0.001	0.652	0.197	0.002	34.53	< 0.001	0.082	1,547
IHS tax returns, 1989	0.329	0.057	< 0.001	0.462	0.126	< 0.001	49.59	< 0.001	0.259	1,547
IHS total annual payroll, 1986	0.309	0.084	< 0.001	0.594	0.173	0.001	24.22	< 0.001	0.169	1,547
IHS total employment, 1986	0.273	0.072	< 0.001	0.552	0.159	0.001	27.94	< 0.001	0.180	1,547
IHS total establishments, 1986	0.232	0.060	< 0.001	0.489	0.140	0.001	30.85	< 0.001	0.220	1,547
IHS wages, 1989	0.440	0.069	< 0.001	0.499	0.137	< 0.001	50.95	< 0.001	0.259	1,547
Log house-price index, 1986	0.070	0.020	0.001	0.058	0.019	0.005	16.15	< 0.001	0.099	760
Logged electricity price, 1990	0.014	0.010	0.165	0.007	0.004	0.073	2.70	0.080	0.008	1,517
IHS public-supply water withdrawals, 1985	0.209	0.065	0.003	0.497	0.127	< 0.001	26.62	< 0.001	0.181	1,547

Notes: Each row is a separate early-level regression. The two alternative exposures—negative log distance to a long-haul fiber node and the 1980 urban-college share—enter jointly after each is standardized in the full exposure file. Controls, cohort restrictions, outcome transformations, state fixed effects, and state-clustered inference match the corresponding preferred-instrument balance panel. No early economic outcome is included as a control in its own balance regression. In this panel, 11 of 12 joint tests reject at 5 percent.

Table C.6: Alternative-exposure pre-1995 placebo changes: Early controls, economic activity

Outcome	Year	2SLS b	(SE)	First-stage F	Conventional p	AR p	N
Total employment	1986	0.081	(0.030)	18.16	0.011	0.072	1,547
	1987	0.061	(0.027)	18.25	0.032	0.185	1,547
	1988	0.055	(0.026)	18.25	0.041	0.191	1,547
	1989	0.048	(0.027)	18.25	0.084	0.157	1,547
	1990	0.049	(0.026)	18.25	0.070	0.152	1,547
	1991	0.043	(0.026)	18.25	0.099	0.175	1,547
	1992	0.040	(0.022)	18.25	0.080	0.237	1,547
	1993	0.029	(0.016)	18.25	0.075	0.281	1,547
	1994	0.009	(0.006)	18.25	0.106	0.361	1,547
Construction employment	1986	0.065	(0.051)	18.25	0.212	0.392	1,547
	1987	0.032	(0.058)	18.25	0.582	0.553	1,547
	1988	-0.012	(0.056)	18.25	0.826	0.870	1,547
	1989	0.005	(0.048)	18.25	0.921	0.467	1,547
	1990	-0.143	(0.055)	18.25	0.013	0.120	1,547
	1991	-0.241	(0.053)	18.25	$<$ 0.001	0.013	1,547
	1992	0.043	(0.043)	18.25	0.319	0.145	1,547
	1993	0.022	(0.039)	18.25	0.575	0.673	1,547
	1994	-0.076	(0.037)	18.25	0.045	0.230	1,547
Establishments	1986	0.023	(0.007)	17.06	0.001	0.003	1,547
	1987	0.018	(0.006)	18.25	0.008	0.020	1,547
	1988	0.018	(0.006)	18.25	0.002	0.010	1,547
	1989	0.015	(0.004)	18.25	0.001	0.012	1,547
	1990	0.013	(0.004)	18.25	0.001	0.016	1,547
	1991	0.012	(0.004)	18.25	0.002	0.019	1,547
	1992	0.007	(0.003)	18.25	0.035	0.028	1,547
	1993	0.005	(0.003)	18.25	0.088	0.123	1,547
	1994	0.003	(0.002)	18.25	0.093	0.354	1,547
Annual payroll	1986	0.104	(0.040)	18.09	0.013	0.112	1,547
	1987	0.053	(0.033)	18.25	0.112	0.292	1,547
	1988	0.049	(0.034)	18.25	0.155	0.271	1,547
	1989	0.047	(0.035)	18.25	0.191	0.191	1,547
	1990	0.047	(0.035)	18.25	0.187	0.206	1,547
	1991	0.046	(0.034)	18.25	0.179	0.185	1,547
	1992	0.044	(0.029)	18.25	0.138	0.296	1,547
	1993	0.038	(0.023)	18.25	0.100	0.335	1,547
	1994	0.010	(0.008)	18.25	0.212	0.456	1,547

Notes: Each row is a separate pre-expansion placebo regression using the fixed 1995–2020 change in IHS cumulative reported annualized data-center revenue as the endogenous treatment. The excluded variables are F_c and U_c entered jointly; multiplying both by the nonzero 2020 value of G_t^{ROW} only rescales their cross-sectional span. Outcome changes are relative to 1995 and use the same transformations and outcome-specific common samples as the preferred-instrument placebo figures. The six stated early economic levels enter as controls, except that an outcome's own early level is omitted from that level's placebo regression. All specifications include state fixed effects and state-clustered inference. The AR column is the joint reduced-form Anderson–Rubin test. In this panel, 8 displayed placebo tests reject at 5 percent by AR inference.

Table C.7: Alternative-exposure pre-1995 placebo changes: Early controls, income, housing, electricity, and water

Outcome	Year	2SLS b	(SE)	First-stage F	Conventional p	AR p	N
Tax returns	1989	0.012	(0.006)	18.15	0.054	0.092	1,546
	1990	-0.020	(0.019)	18.25	0.317	0.044	1,546
	1991	0.002	(0.006)	18.25	0.790	0.056	1,546
	1992	-0.000	(0.006)	18.25	0.936	0.070	1,546
	1993	-0.003	(0.006)	18.25	0.566	0.136	1,546
	1994	-0.004	(0.005)	18.25	0.376	0.203	1,546
Adjusted gross income	1989	0.013	(0.008)	18.38	0.113	0.037	1,546
	1990	-0.024	(0.024)	18.25	0.311	0.014	1,546
	1991	0.003	(0.007)	18.25	0.687	0.053	1,546
	1992	0.002	(0.007)	18.25	0.748	0.059	1,546
	1993	-0.004	(0.006)	18.25	0.477	0.092	1,546
	1994	-0.007	(0.005)	18.25	0.217	0.152	1,546
Wages and salaries	1989	0.009	(0.006)	18.60	0.166	0.013	1,546
	1990	-0.029	(0.024)	18.25	0.234	0.008	1,546
	1991	-0.001	(0.007)	18.25	0.935	0.015	1,546
	1992	-0.001	(0.006)	18.25	0.815	0.029	1,546
	1993	-0.004	(0.006)	18.25	0.493	0.078	1,546
	1994	-0.007	(0.006)	18.25	0.249	0.128	1,546
House-price-index growth	1986	-0.008	(0.006)	21.13	0.175	0.134	739
	1987	-0.006	(0.006)	21.13	0.296	0.193	739
	1988	-0.000	(0.005)	21.13	0.955	0.586	739
	1989	0.005	(0.005)	21.13	0.307	0.544	739
	1990	0.007	(0.005)	21.13	0.147	0.404	739
	1991	0.007	(0.005)	21.13	0.151	0.409	739
	1992	0.005	(0.004)	21.13	0.160	0.476	739
	1993	0.003	(0.003)	21.13	0.287	0.604	739
	1994	0.002	(0.002)	21.13	0.204	0.422	739
Electricity price	1990	0.005	(0.004)	13.95	0.211	0.505	1,492
	1991	0.010	(0.006)	13.95	0.120	0.374	1,492
	1992	0.006	(0.004)	13.95	0.186	0.456	1,492
	1993	0.004	(0.001)	13.95	0.006	0.043	1,492
	1994	0.001	(0.002)	13.95	0.421	0.400	1,492
Public-supply water withdrawals	1985	0.036	(0.028)	18.25	0.211	0.465	1,547
	1990	0.049	(0.018)	18.25	0.011	0.062	1,547

Notes: Each row is a separate pre-expansion placebo regression using the fixed 1995–2020 change in IHS cumulative reported annualized data-center revenue as the endogenous treatment. The excluded variables are F_c and U_c entered jointly; multiplying both by the nonzero 2020 value of G_t^{ROW} only rescales their cross-sectional span. Outcome changes are relative to 1995 and use the same transformations and outcome-specific common samples as the preferred-instrument placebo figures. The six stated early economic levels enter as controls, except that an outcome's own early level is omitted from that level's placebo regression. All specifications include state fixed effects and state-clustered inference. The AR column is the joint reduced-form Anderson–Rubin test. In this panel, 8 displayed placebo tests reject at 5 percent by AR inference.

Table C.8: Alternative-exposure pre-1995 placebo changes: Early + terrain controls, economic activity

Outcome	Year	2SLS b	(SE)	First-stage F	Conventional p	AR p	N
Total employment	1986	0.080	(0.030)	17.61	0.012	0.074	1,547
	1987	0.061	(0.027)	17.64	0.031	0.179	1,547
	1988	0.055	(0.026)	17.64	0.040	0.189	1,547
	1989	0.048	(0.027)	17.64	0.081	0.179	1,547
	1990	0.049	(0.026)	17.64	0.068	0.176	1,547
	1991	0.044	(0.026)	17.64	0.097	0.210	1,547
	1992	0.041	(0.022)	17.64	0.077	0.239	1,547
	1993	0.030	(0.016)	17.64	0.068	0.268	1,547
	1994	0.010	(0.006)	17.64	0.116	0.371	1,547
Construction employment	1986	0.066	(0.051)	17.64	0.203	0.441	1,547
	1987	0.034	(0.057)	17.64	0.554	0.794	1,547
	1988	-0.009	(0.054)	17.64	0.863	0.985	1,547
	1989	0.008	(0.047)	17.64	0.867	0.717	1,547
	1990	-0.138	(0.054)	17.64	0.014	0.122	1,547
	1991	-0.239	(0.053)	17.64	$<$ 0.001	0.007	1,547
	1992	0.047	(0.044)	17.64	0.288	0.268	1,547
	1993	0.025	(0.040)	17.64	0.531	0.833	1,547
	1994	-0.076	(0.036)	17.64	0.043	0.226	1,547
Establishments	1986	0.023	(0.007)	16.63	$<$ 0.001	0.004	1,547
	1987	0.018	(0.006)	17.64	0.006	0.019	1,547
	1988	0.018	(0.005)	17.64	0.002	0.011	1,547
	1989	0.015	(0.004)	17.64	$<$ 0.001	0.013	1,547
	1990	0.013	(0.004)	17.64	$<$ 0.001	0.014	1,547
	1991	0.012	(0.004)	17.64	0.002	0.026	1,547
	1992	0.007	(0.003)	17.64	0.034	0.041	1,547
	1993	0.005	(0.003)	17.64	0.084	0.135	1,547
	1994	0.003	(0.002)	17.64	0.095	0.355	1,547
Annual payroll	1986	0.104	(0.040)	17.50	0.013	0.111	1,547
	1987	0.054	(0.033)	17.64	0.109	0.284	1,547
	1988	0.050	(0.034)	17.64	0.148	0.274	1,547
	1989	0.047	(0.035)	17.64	0.186	0.220	1,547
	1990	0.047	(0.035)	17.64	0.182	0.235	1,547
	1991	0.047	(0.034)	17.64	0.175	0.225	1,547
	1992	0.046	(0.030)	17.64	0.131	0.291	1,547
	1993	0.039	(0.023)	17.64	0.093	0.322	1,547
	1994	0.011	(0.009)	17.64	0.215	0.445	1,547

Notes: Each row is a separate pre-expansion placebo regression using the fixed 1995–2020 change in IHS cumulative reported annualized data-center revenue as the endogenous treatment. The excluded variables are F_c and U_c entered jointly; multiplying both by the nonzero 2020 value of G_t^{ROW} only rescales their cross-sectional span. Outcome changes are relative to 1995 and use the same transformations and outcome-specific common samples as the preferred-instrument placebo figures. The six stated early economic levels enter as controls, except that an outcome's own early level is omitted from that level's placebo regression. All specifications include state fixed effects and state-clustered inference. The AR column is the joint reduced-form Anderson–Rubin test. In this panel, 8 displayed placebo tests reject at 5 percent by AR inference.

Table C.9: Alternative-exposure pre-1995 placebo changes: Early + terrain controls, income, housing, electricity, and water

Outcome	Year	2SLS b	(SE)	First-stage F	Conventional p	AR p	N
Tax returns	1989	0.011	(0.006)	17.58	0.057	0.080	1,546
	1990	-0.019	(0.019)	17.63	0.316	0.043	1,546
	1991	0.002	(0.007)	17.63	0.773	0.047	1,546
	1992	-0.000	(0.006)	17.63	0.951	0.053	1,546
	1993	-0.003	(0.006)	17.63	0.571	0.102	1,546
	1994	-0.004	(0.005)	17.63	0.370	0.159	1,546
Adjusted gross income	1989	0.013	(0.008)	17.75	0.118	0.018	1,546
	1990	-0.024	(0.023)	17.63	0.307	0.007	1,546
	1991	0.003	(0.007)	17.63	0.688	0.025	1,546
	1992	0.002	(0.007)	17.63	0.757	0.023	1,546
	1993	-0.004	(0.006)	17.63	0.468	0.037	1,546
	1994	-0.007	(0.005)	17.63	0.207	0.082	1,546
Wages and salaries	1989	0.009	(0.006)	17.99	0.170	0.009	1,546
	1990	-0.028	(0.023)	17.63	0.236	0.006	1,546
	1991	-0.000	(0.007)	17.63	0.973	0.010	1,546
	1992	-0.001	(0.007)	17.63	0.838	0.018	1,546
	1993	-0.004	(0.006)	17.63	0.498	0.043	1,546
	1994	-0.007	(0.006)	17.63	0.245	0.085	1,546
House-price-index growth	1986	-0.008	(0.006)	20.38	0.194	0.152	739
	1987	-0.006	(0.006)	20.38	0.315	0.233	739
	1988	-0.000	(0.005)	20.38	0.929	0.568	739
	1989	0.005	(0.005)	20.38	0.326	0.527	739
	1990	0.007	(0.005)	20.38	0.159	0.354	739
	1991	0.006	(0.005)	20.38	0.168	0.342	739
	1992	0.005	(0.004)	20.38	0.182	0.431	739
	1993	0.003	(0.003)	20.38	0.320	0.565	739
	1994	0.003	(0.002)	20.38	0.200	0.482	739
Electricity price	1990	0.005	(0.004)	13.60	0.212	0.425	1,492
	1991	0.010	(0.006)	13.60	0.119	0.341	1,492
	1992	0.005	(0.004)	13.60	0.181	0.405	1,492
	1993	0.004	(0.001)	13.60	0.006	0.022	1,492
	1994	0.001	(0.002)	13.60	0.439	0.325	1,492
Public-supply water withdrawals	1985	0.036	(0.028)	17.64	0.213	0.462	1,547
	1990	0.049	(0.018)	17.64	0.011	0.062	1,547

Notes: Each row is a separate pre-expansion placebo regression using the fixed 1995–2020 change in IHS cumulative reported annualized data-center revenue as the endogenous treatment. The excluded variables are F_c and U_c entered jointly; multiplying both by the nonzero 2020 value of G_t^{ROW} only rescales their cross-sectional span. Outcome changes are relative to 1995 and use the same transformations and outcome-specific common samples as the preferred-instrument placebo figures. The six stated early economic levels enter as controls, except that an outcome's own early level is omitted from that level's placebo regression. All specifications include state fixed effects and state-clustered inference. The AR column is the joint reduced-form Anderson–Rubin test. In this panel, 13 displayed placebo tests reject at 5 percent by AR inference.

D Ordinary least-squares comparisons

This appendix reports the OLS counterparts to the paper's two main IV tables. These regressions hold fixed each IV design's treatment, outcome transformation, controls, fixed effects, clustering, and estimation sample, but do not instrument cumulative data-center revenue. They are descriptive conditional associations and need not identify the causal effects recovered by the historical instruments.

Table D.1: OLS counterpart to the pooled annual specification

Outcome	End	OLS estimate	(SE)	p	Implied effect at mean ΔD	N
Total employment	2023	0.031***	(0.007)	< 0.001	0.7%	45,733
Construction employment	2023	0.256***	(0.038)	< 0.001	5.7%	45,733
Establishments	2023	0.033***	(0.006)	< 0.001	0.7%	45,733
Annual payroll	2023	0.028***	(0.006)	< 0.001	0.6%	45,733
Tax returns	2022	0.033***	(0.007)	< 0.001	0.7%	44,154
Adjusted gross income	2022	0.032***	(0.007)	< 0.001	0.7%	44,154
Wages and salaries	2022	0.030***	(0.007)	< 0.001	0.7%	44,154
House-price index	2023	0.009***	(0.003)	0.002	0.2%	41,248
Electricity-price elasticity	2023	-0.00205	(0.00423)	0.631	0.0%	45,079
Public-supply water withdrawals	2020	0.022***	(0.004)	< 0.001	0.3%	33,117

Notes: Every row uses the same annual sample, treatment, outcome transformation, county fixed effects, state-by-year fixed effects, and state-clustered inference as Table 1; cumulative data-center revenue is not instrumented. The implied-effect column evaluates each coefficient at the outcome sample's mean treatment change. Electricity prices enter in natural logs. $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$.

Reading the first coefficient: The estimate of 0.031 implies that 10 percent more cumulative data-center revenue is associated with approximately 0.3 percent more total employment. At the mean treatment change of 0.217 IHS points, the implied association is 0.7 percent.

Table D.2: OLS counterpart to the latest-endpoint long differences

Outcome	End	OLS estimate	(SE)	p	Implied effect at mean ΔD	N
Total employment	2023	0.028**	(0.013)	0.036	0.6%	1,547
Construction employment	2023	0.048	(0.031)	0.176	1.0%	1,547
Establishments	2023	0.018*	(0.009)	0.054	0.4%	1,547
Annual payroll	2023	0.056***	(0.014)	< 0.001	1.2%	1,547
Tax returns	2022	0.009	(0.010)	0.391	0.2%	1,547
Adjusted gross income	2022	0.030***	(0.011)	0.007	0.6%	1,547
Wages and salaries	2022	0.033**	(0.012)	0.011	0.7%	1,547
House-price-index growth	2023	0.084***	(0.011)	< 0.001	2.5 pp	1,100
Electricity-price elasticity	2023	-0.00309	(0.00424)	0.471	-0.1%	1,498
Public-supply water withdrawals	2020	0.019	(0.018)	0.313	0.4%	1,547

Notes: Every row uses the same 1995-to-endpoint change, outcome-specific sample, Early + terrain controls, state fixed effects, and state-clustered inference as Table 2; cumulative data-center revenue is not instrumented. The implied-effect column evaluates each coefficient at the outcome sample's mean treatment change. Electricity prices enter in natural logs. $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$.

Reading the first coefficient: The estimate of 0.028 implies that 10 percent more cumulative data-center revenue is associated with approximately 0.3 percent more total employment between 1995 and 2023. At the mean treatment change of 0.214 IHS points, the implied association is 0.6 percent.

E Historical-corridor 2SLS results

The following tables report all ten outcomes at the five pre-specified endpoints under both control sets. Each cell includes the coefficient, state-clustered standard error, the outcome-specific first-stage statistic, and the inferential p-value. Whenever $F < 10$, that p-value is the joint Anderson–Rubin p-value and the stars use the same weak-IV-robust test. Thus, conventional 2SLS significance is never used to support a weakly identified cell.

Table E.1: Joint-instrument 2SLS estimates: Early controls

Outcome	2000	2005	2010	2015	2020
Total employment	0.457 (0.382) AR $p = 0.319$ $F = 1.85$	0.381 (0.302) AR $p = 0.298$ $F = 2.14$	0.637 (0.373) AR $p = 0.102$ $F = 4.73$	0.304** (0.125) $p = 0.020$ $F = 10.44$	0.203* (0.113) $p = 0.081$ $F = 13.74$
Construction employment	1.225 (1.203) AR $p = 0.523$ $F = 1.85$	1.016 (1.821) AR $p = 0.550$ $F = 2.14$	2.968* (1.647) AR $p = 0.099$ $F = 4.73$	1.110* (0.557) $p = 0.054$ $F = 10.44$	0.563 (0.343) $p = 0.109$ $F = 13.74$
Establishments	0.056 (0.084) AR $p = 0.619$ $F = 1.85$	0.193 (0.177) AR $p = 0.188$ $F = 2.14$	0.224* (0.132) AR $p = 0.060$ $F = 4.73$	0.191* (0.096) $p = 0.053$ $F = 10.44$	0.187** (0.088) $p = 0.039$ $F = 13.74$
Annual payroll	0.805 (0.585) AR $p = 0.193$ $F = 1.85$	0.703 (0.473) AR $p = 0.183$ $F = 2.14$	0.757** (0.466) AR $p = 0.030$ $F = 4.73$	0.451** (0.169) $p = 0.011$ $F = 10.44$	0.342** (0.149) $p = 0.027$ $F = 13.74$
Tax returns	0.141*** (0.088) AR $p = 0.008$ $F = 1.85$	0.140* (0.121) AR $p = 0.075$ $F = 2.14$	-0.139** (0.312) AR $p = 0.027$ $F = 4.70$	0.078 (0.077) $p = 0.317$ $F = 10.44$	0.065 (0.074) $p = 0.382$ $F = 13.74$
Adjusted gross income	0.347*** (0.200) AR $p = 0.003$ $F = 1.85$	0.300*** (0.198) AR $p = 0.004$ $F = 2.14$	-0.579 (0.698) AR $p = 0.586$ $F = 4.70$	0.182** (0.085) $p = 0.040$ $F = 10.44$	0.142 (0.093) $p = 0.135$ $F = 13.74$
Wages and salaries	0.355*** (0.196) AR $p = 0.001$ $F = 1.85$	0.297** (0.171) AR $p = 0.019$ $F = 2.14$	-0.351** (0.660) AR $p = 0.015$ $F = 4.70$	0.179** (0.085) $p = 0.042$ $F = 10.44$	0.166* (0.092) $p = 0.080$ $F = 13.74$
House-price-index growth	0.183*** (0.129) AR $p = 0.004$ $F = 1.70$	0.441*** (0.267) AR $p = 0.007$ $F = 1.71$	0.151 (0.100) AR $p = 0.147$ $F = 3.14$	0.232** (0.092) AR $p = 0.010$ $F = 6.79$	0.319*** (0.109) AR $p = 0.000$ $F = 8.81$
Electricity-price elasticity	-0.009 (0.045) AR $p = 0.861$ $F = 2.19$	-0.010 (0.085) AR $p = 0.754$ $F = 2.47$	0.022 (0.109) AR $p = 0.977$ $F = 4.91$	0.024 (0.080) AR $p = 0.880$ $F = 9.92$	0.061 (0.069) $p = 0.382$ $F = 11.01$
Public-supply water withdrawals	0.513 (0.451) AR $p = 0.256$ $F = 1.85$	0.519 (0.430) AR $p = 0.276$ $F = 2.14$	0.461 (0.305) AR $p = 0.163$ $F = 4.73$	0.330* (0.183) $p = 0.081$ $F = 10.44$	0.306* (0.156) $p = 0.058$ $F = 13.74$

Notes: Each cell reports the 2SLS coefficient, state-clustered standard error, relevant inferential p-value, and the outcome-specific first-stage statistic. When $F < 10$, stars and inference use the joint Anderson–Rubin test of a zero structural effect; when $F \geq 10$, they use the conventional state-clustered 2SLS test. $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$. All outcomes are reported; electricity price is in natural logs, so its coefficient has the same approximate elasticity interpretation as the other positive-valued outcomes. To read the first coefficient, at positive values where IHS approximates a log, 0.457 implies that a 10 percent increase in cumulative data-center revenue raises total employment by approximately 4.6 percent between 1995 and 2000. This early estimate is weakly identified ($F = 1.85$) and imprecise under Anderson–Rubin inference.

Table E.2: Joint-instrument 2SLS estimates: Early + terrain controls

Outcome	2000	2005	2010	2015	2020
Total employment	0.432 (0.382) AR $p = 0.417$ $F = 1.79$	0.416 (0.324) AR $p = 0.308$ $F = 2.05$	0.729 (0.417) AR $p = 0.116$ $F = 4.33$	0.344* (0.149) AR $p = 0.060$ $F = 8.05$	0.221* (0.121) $p = 0.076$ $F = 11.61$
Construction employment	0.967 (1.134) AR $p = 0.600$ $F = 1.79$	0.925 (1.875) AR $p = 0.430$ $F = 2.05$	3.110 (1.789) AR $p = 0.106$ $F = 4.33$	1.273** (0.565) AR $p = 0.036$ $F = 8.05$	0.614* (0.345) $p = 0.083$ $F = 11.61$
Establishments	0.065 (0.083) AR $p = 0.654$ $F = 1.79$	0.226 (0.184) AR $p = 0.206$ $F = 2.05$	0.255* (0.141) AR $p = 0.065$ $F = 4.33$	0.223** (0.102) AR $p = 0.027$ $F = 8.05$	0.208** (0.091) $p = 0.028$ $F = 11.61$
Annual payroll	0.780 (0.589) AR $p = 0.266$ $F = 1.79$	0.762 (0.511) AR $p = 0.198$ $F = 2.05$	0.857** (0.487) AR $p = 0.041$ $F = 4.33$	0.514** (0.199) AR $p = 0.029$ $F = 8.05$	0.371** (0.166) $p = 0.031$ $F = 11.61$
Tax returns	0.159** (0.097) AR $p = 0.011$ $F = 1.79$	0.178* (0.126) AR $p = 0.064$ $F = 2.05$	-0.080** (0.295) AR $p = 0.023$ $F = 4.28$	0.115* (0.080) AR $p = 0.058$ $F = 8.05$	0.089 (0.074) $p = 0.240$ $F = 11.61$
Adjusted gross income	0.369*** (0.213) AR $p = 0.003$ $F = 1.79$	0.360*** (0.212) AR $p = 0.004$ $F = 2.05$	-0.560 (0.731) AR $p = 0.647$ $F = 4.28$	0.230*** (0.091) AR $p = 0.004$ $F = 8.05$	0.177* (0.094) $p = 0.067$ $F = 11.61$
Wages and salaries	0.364*** (0.203) AR $p = 0.001$ $F = 1.79$	0.334** (0.181) AR $p = 0.014$ $F = 2.05$	-0.265** (0.615) AR $p = 0.012$ $F = 4.28$	0.221*** (0.089) AR $p = 0.008$ $F = 8.05$	0.195** (0.094) $p = 0.044$ $F = 11.61$
House-price-index growth	0.161*** (0.121) AR $p = 0.002$ $F = 1.63$	0.410*** (0.252) AR $p = 0.003$ $F = 1.64$	0.178** (0.120) AR $p = 0.022$ $F = 2.85$	0.245*** (0.105) AR $p = 0.002$ $F = 5.55$	0.325*** (0.119) AR $p = 0.000$ $F = 7.82$
Electricity-price elasticity	-0.010 (0.040) AR $p = 0.783$ $F = 2.13$	-0.002 (0.074) AR $p = 0.709$ $F = 2.37$	0.021 (0.100) AR $p = 0.921$ $F = 4.57$	0.018 (0.075) AR $p = 0.786$ $F = 7.76$	0.056 (0.062) AR $p = 0.185$ $F = 9.10$
Public-supply water withdrawals	0.494 (0.451) AR $p = 0.253$ $F = 1.79$	0.498 (0.434) AR $p = 0.280$ $F = 2.05$	0.465 (0.320) AR $p = 0.160$ $F = 4.33$	0.335 (0.198) AR $p = 0.140$ $F = 8.05$	0.311* (0.163) $p = 0.065$ $F = 11.61$

Notes: Each cell reports the 2SLS coefficient, state-clustered standard error, relevant inferential p-value, and the outcome-specific first-stage statistic. When $F < 10$, stars and inference use the joint Anderson–Rubin test of a zero structural effect; when $F \geq 10$, they use the conventional state-clustered 2SLS test. $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$. All outcomes are reported; electricity price is in natural logs, so its coefficient has the same approximate elasticity interpretation as the other positive-valued outcomes. To read the first coefficient, at positive values where IHS approximates a log, 0.432 implies that a 10 percent increase in cumulative data-center revenue raises total employment by approximately 4.3 percent between 1995 and 2000. This early estimate is weakly identified ($F = 1.79$) and imprecise under Anderson–Rubin inference.

F Early-level balance results

The following four panels report the complete 48-test balance battery for 1860 and 1870 historical-population cohorts, with and without direct terrain controls. No joint test rejects at 5 percent.

Table F.1: Joint early-level balance: 1860 cohort, historical population and geography

Early outcome	Access-node b	Access-node SE	Access-node p	Terrain b	Terrain SE	Terrain p	Joint F	Joint p	Partial R^2	N
IHS adjusted gross income, 1989	-0.002	0.040	0.957	0.018	0.138	0.895	0.01	0.990	0.000	1,270
IHS construction employment, 1986	0.003	0.065	0.965	-0.047	0.172	0.787	0.04	0.962	0.000	1,270
IHS construction establishments, 1986	0.005	0.045	0.910	0.029	0.117	0.806	0.03	0.966	0.000	1,270
IHS construction payroll, 1986	-0.002	0.083	0.985	-0.072	0.203	0.725	0.06	0.939	0.000	1,270
IHS tax returns, 1989	0.000	0.037	0.995	0.022	0.122	0.856	0.02	0.984	0.000	1,270
IHS total annual payroll, 1986	-0.021	0.058	0.719	-0.013	0.152	0.932	0.07	0.931	0.000	1,270
IHS total employment, 1986	-0.011	0.054	0.834	-0.006	0.134	0.965	0.02	0.976	0.000	1,270
IHS total establishments, 1986	0.005	0.045	0.917	0.014	0.109	0.900	0.01	0.987	0.000	1,270
IHS wages, 1989	-0.007	0.040	0.870	0.010	0.152	0.947	0.02	0.985	0.000	1,270
Log house-price index, 1986	-0.013	0.011	0.255	0.039	0.028	0.197	1.72	0.251	0.004	663
Logged electricity price, 1990	0.00143	0.00479	0.767	0.01798	0.01547	0.253	0.86	0.433	0.003	1,242
IHS public-supply water withdrawals, 1985	0.002	0.046	0.973	0.083	0.106	0.435	0.32	0.729	0.001	1,270

Notes: Each row is a separate balance regression. The two instruments are the rail access-node interaction and terrain-route proximity, tested jointly with state-clustered standard errors. Controls include cohort log population, geography, state fixed effects, and both lower-order rail-coverage and access-node-geometry components; terrain panels additionally include direct elevation and slope controls. No early economic outcome is included as a control in its own balance regression. Across all four panels, none of the 48 joint tests rejects at 5 percent. To read the first coefficient, a one-standard-deviation increase in the access-node measure is associated with a 0.002-IHS-point decrease—approximately 0.2 percent—in 1989 adjusted gross income; the estimate is essentially zero ($p = 0.957$).

Table F.2: Joint early-level balance: 1860 cohort, add direct terrain controls

Early outcome	Access-node b	Access-node SE	Access-node p	Terrain b	Terrain SE	Terrain p	Joint F	Joint p	Partial R^2	N
IHS adjusted gross income, 1989	0.002	0.041	0.967	0.022	0.133	0.869	0.01	0.986	0.000	1,270
IHS construction employment, 1986	0.007	0.069	0.920	-0.011	0.175	0.952	0.01	0.993	0.000	1,270
IHS construction establishments, 1986	0.010	0.047	0.838	0.037	0.115	0.748	0.07	0.934	0.000	1,270
IHS construction payroll, 1986	0.001	0.088	0.993	-0.015	0.203	0.942	0.00	0.997	0.000	1,270
IHS tax returns, 1989	0.003	0.038	0.935	0.026	0.118	0.829	0.03	0.973	0.000	1,270
IHS total annual payroll, 1986	-0.014	0.059	0.817	-0.040	0.147	0.788	0.07	0.937	0.000	1,270
IHS total employment, 1986	-0.004	0.055	0.937	-0.029	0.132	0.827	0.03	0.972	0.000	1,270
IHS total establishments, 1986	0.008	0.046	0.865	0.020	0.109	0.853	0.03	0.968	0.000	1,270
IHS wages, 1989	-0.003	0.041	0.950	0.012	0.148	0.935	0.01	0.995	0.000	1,270
Log house-price index, 1986	-0.011	0.011	0.330	0.020	0.030	0.507	0.79	0.465	0.002	663
Logged electricity price, 1990	0.00056	0.00461	0.904	0.02933	0.01329	0.034	2.78	0.076	0.006	1,242
IHS public-supply water withdrawals, 1985	0.007	0.048	0.887	0.070	0.106	0.511	0.24	0.789	0.001	1,270

Notes: Each row is a separate balance regression. The two instruments are the rail access-node interaction and terrain-route proximity, tested jointly with state-clustered standard errors. Controls include cohort log population, geography, state fixed effects, and both lower-order rail-coverage and access-node-geometry components; terrain panels additionally include direct elevation and slope controls. No early economic outcome is included as a control in its own balance regression. Across all four panels, none of the 48 joint tests rejects at 5 percent. To read the first coefficient, a one-standard-deviation increase in the access-node measure is associated with a 0.002-IHS-point increase—approximately 0.2 percent—in 1989 adjusted gross income; the estimate is essentially zero ($p = 0.967$).

Table F.3: Joint early-level balance: 1870 cohort, historical population and geography

Early outcome	Access-node b	Access-node SE	Access-node p	Terrain b	Terrain SE	Terrain p	Joint F	Joint p	Partial R^2	N
IHS adjusted gross income, 1989	-0.032	0.037	0.395	-0.075	0.141	0.597	0.49	0.615	0.002	1,547
IHS construction employment, 1986	-0.046	0.060	0.442	-0.246	0.206	0.241	0.94	0.400	0.003	1,547
IHS construction establishments, 1986	-0.014	0.043	0.750	-0.089	0.132	0.504	0.27	0.767	0.001	1,547
IHS construction payroll, 1986	-0.068	0.075	0.369	-0.351	0.249	0.167	1.29	0.288	0.004	1,547
IHS tax returns, 1989	-0.017	0.033	0.624	-0.056	0.125	0.657	0.22	0.806	0.001	1,547
IHS total annual payroll, 1986	-0.060	0.054	0.274	-0.087	0.161	0.594	0.75	0.479	0.002	1,547
IHS total employment, 1986	-0.043	0.050	0.388	-0.076	0.140	0.593	0.52	0.599	0.002	1,547
IHS total establishments, 1986	-0.015	0.041	0.715	-0.050	0.111	0.658	0.16	0.853	0.001	1,547
IHS wages, 1989	-0.029	0.036	0.427	-0.088	0.156	0.575	0.48	0.625	0.002	1,547
Log house-price index, 1986	-0.013	0.010	0.220	0.036	0.027	0.200	2.19	0.126	0.004	760
Logged electricity price, 1990	0.00211	0.00433	0.628	0.01231	0.01321	0.357	0.63	0.536	0.002	1,517
IHS public-supply water withdrawals, 1985	-0.019	0.043	0.654	0.006	0.120	0.962	0.10	0.903	0.000	1,547

Notes: Each row is a separate balance regression. The two instruments are the rail access-node interaction and terrain-route proximity, tested jointly with state-clustered standard errors. Controls include cohort log population, geography, state fixed effects, and both lower-order rail-coverage and access-node-geometry components; terrain panels additionally include direct elevation and slope controls. No early economic outcome is included as a control in its own balance regression. Across all four panels, none of the 48 joint tests rejects at 5 percent. To read the first coefficient, a one-standard-deviation increase in the access-node measure is associated with a 0.032-IHS-point decrease—approximately 3.2 percent—in 1989 adjusted gross income; the estimate is statistically indistinguishable from zero ($p = 0.395$).

Table F.4: Joint early-level balance: 1870 cohort, add direct terrain controls

Early outcome	Access-node b	Access-node SE	Access-node p	Terrain b	Terrain SE	Terrain p	Joint F	Joint p	Partial R^2	N
IHS adjusted gross income, 1989	-0.029	0.037	0.447	-0.077	0.139	0.584	0.42	0.659	0.002	1,547
IHS construction employment, 1986	-0.045	0.063	0.475	-0.213	0.207	0.308	0.77	0.472	0.003	1,547
IHS construction establishments, 1986	-0.011	0.044	0.810	-0.081	0.129	0.534	0.22	0.805	0.001	1,547
IHS construction payroll, 1986	-0.069	0.079	0.384	-0.303	0.250	0.234	1.07	0.354	0.003	1,547
IHS tax returns, 1989	-0.014	0.034	0.689	-0.058	0.124	0.639	0.19	0.831	0.001	1,547
IHS total annual payroll, 1986	-0.054	0.054	0.328	-0.102	0.158	0.523	0.67	0.517	0.002	1,547
IHS total employment, 1986	-0.037	0.050	0.462	-0.091	0.138	0.515	0.48	0.625	0.002	1,547
IHS total establishments, 1986	-0.013	0.042	0.758	-0.046	0.109	0.677	0.13	0.878	0.001	1,547
IHS wages, 1989	-0.026	0.037	0.479	-0.090	0.155	0.565	0.41	0.666	0.002	1,547
Log house-price index, 1986	-0.011	0.010	0.290	0.018	0.029	0.533	0.98	0.391	0.002	760
Logged electricity price, 1990	0.00136	0.00410	0.742	0.01806	0.01197	0.140	1.31	0.281	0.003	1,517
IHS public-supply water withdrawals, 1985	-0.014	0.043	0.740	-0.009	0.119	0.939	0.06	0.942	0.000	1,547

Notes: Each row is a separate balance regression. The two instruments are the rail access-node interaction and terrain-route proximity, tested jointly with state-clustered standard errors. Controls include cohort log population, geography, state fixed effects, and both lower-order rail-coverage and access-node-geometry components; terrain panels additionally include direct elevation and slope controls. No early economic outcome is included as a control in its own balance regression. Across all four panels, none of the 48 joint tests rejects at 5 percent. To read the first coefficient, a one-standard-deviation increase in the access-node measure is associated with a 0.029-IHS-point decrease—approximately 2.9 percent—in 1989 adjusted gross income; the estimate is statistically indistinguishable from zero ($p = 0.447$).

G Property-tax revenue

Data. We measure property-tax revenue using the T01 series in the U.S. Census Bureau's Annual Survey of State and Local Government Finances and Census of Governments public-use files. The source panel spans 1995–2023 and combines the full Census of Governments waves in 1997, 2002, 2007, 2012, 2017, and 2022 with the intervening annual survey releases. T01 is nominal property-tax revenue in thousands of dollars. We sum receipts across county, municipal, school-district, special-district, and other local-government units assigned to a county area. The outcome therefore measures property-tax collections by all local governments located within the county; it is not limited to the county government or to revenue ultimately retained by it. Before 2017, the source identifier is a Census county-area code rather than modern FIPS, so we use the existing identifier-only crosswalk to recover county FIPS without importing outcome information. The regressions use the IHS transformation because the early annual files contain recorded zeros.

Pooled shift-share results. Table G.1 applies the pooled historical-corridor shift-share specification to the annual IHS series, the positive-value log series, and the full-Census waves. The annual IHS estimate is positive but imprecise: the coefficient is 0.241, the first-stage statistic is 9.13, and the weak-IV-robust Anderson–Rubin p-value is 0.746. Restricting the outcome to positive observations or to the full-Census waves does not produce statistically robust evidence of an effect.

Table G.1: Property-tax revenue: pooled historical-corridor shift-share

Sample and transformation	Estimate	SE	F	Conventional p	AR p	Mean ΔD	Implied effect	N
Annual, IHS	0.241	0.760	9.13	0.753	0.746	0.217	5.4%	43,743
Annual, log positive values	0.284	0.426	9.11	0.509	0.477	0.217	6.4%	43,476
Census waves, IHS	0.207	0.173	9.67	0.239	0.193	0.208	4.4%	9,424
Census waves, log positive values	0.207	0.173	9.67	0.239	0.193	0.208	4.4%	9,424

Notes: The excluded variable is $s_c^S G_t^S$, the current paper's historical-corridor exposure share interacted with leave-U.S.-out scale-oriented facility growth. All models include county and state-by-year fixed effects and cluster standard errors by state. The treatment D is IHS cumulative annualized county data-center revenue. The annual sample spans 1995–2023; the Census-wave sample uses 1997, 2002, 2007, 2012, 2017, and 2022. Every F is below 10, so stars and substantive conclusions use the weak-IV-robust Anderson–Rubin (AR) test. The implied effect is $100[\exp(\hat{\beta}\,\overline{\Delta D}) - 1]$ and is only an approximate percentage interpretation for IHS outcomes. $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$.

Long differences and source sensitivity. Table G.2 reports selected 1995-to-endpoint long differences. In the preferred 1995–2023 specification, the coefficient is 1.465, but the first-stage statistic is 8.98 and the Anderson–Rubin p-value is 0.159. Excluding recorded zeros reduces that endpoint estimate to 0.224, with an Anderson–Rubin p-value of 0.710. The sensitivity matters because the 1995 baseline is an Annual Survey release with incomplete coverage and 117 recorded zeros.

Table G.2: Property-tax revenue: selected 1995-to-endpoint long differences

		Early controls, IHS			Preferred controls, IHS			Preferred, log positive		
End	Release	Estimate (SE)	F	AR p	Estimate (SE)	F	AR p	Estimate (SE)	F	AR p
2000	Annual	3.719 (2.968)	1.87	0.237	3.832 (3.085)	1.78	0.227	0.487 (0.991)	1.73	0.259
2005	Annual	3.482* (2.648)	1.93	0.085	3.484 (2.713)	1.78	0.181	0.520 (0.991)	1.72	0.842
2010	Annual	3.151 (1.988)	3.20	0.133	3.443 (2.351)	2.56	0.193	0.837 (0.933)	2.54	0.509
2012	Census	2.745* (1.607)	3.21	0.078	2.879 (1.809)	2.62	0.128	0.635 (0.751)	2.49	0.631
2015	Annual	1.407 (0.734)	8.18	0.106	1.418 (0.820)	5.71	0.250	0.167 (0.433)	5.50	0.895
2017	Census	1.779* (0.887)	8.12	0.080	1.857 (0.995)	5.90	0.130	0.428 (0.421)	5.74	0.604
2020	Annual	1.429* (0.706)	10.90	0.119	1.534 (0.790)	8.60	0.152	0.271 (0.330)	8.41	0.516
2022	Census	1.538** (0.740)	11.42	0.086	1.590 (0.806)	9.21	0.128	0.409 (0.380)	8.88	0.551
2023	Annual	1.377* (0.695)	11.40	0.127	1.465 (0.766)	8.98	0.159	0.224 (0.336)	8.79	0.710

Notes: Each row is a separate cross-sectional 2SLS regression of the 1995-to-endpoint property-tax change on the corresponding change in IHS cumulative data-center revenue. The two excluded variables are the historical network-access-node and terrain-predicted-route measures. All specifications include state fixed effects, historical geography, 1870 population, the lower-order rail components, and 1986/1989 economic levels; the preferred specification adds direct terrain controls. Standard errors are clustered by state. Stars use AR inference whenever $F < 10$ and conventional clustered inference otherwise. AR p is shown in every row for transparency. $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$.

Table G.3 therefore uses the 1997 full Census of Governments as the baseline. The 2012 estimate is positive and rejects zero under weak-IV-robust inference, but the effect does not persist at the 2022 Census endpoint or the 2023 Annual Survey check. Taken together, the estimates are generally positive but imprecise.

Table G.3: Sensitivity: 1997 full-Census baseline

		IHS change				Winsorized IHS change			
End	Release	Estimate (SE)	F	Conventional p	AR p	Estimate (SE)	F	Conventional p	AR p
2002	Census	0.095 (0.129)	2.86	0.467	0.474	0.117 (0.127)	2.86	0.362	0.265
2007	Census	0.190 (0.165)	2.53	0.259	0.340	0.209 (0.170)	2.53	0.227	0.236
2012	Census	0.255*** (0.154)	3.94	0.106	0.007	0.247*** (0.147)	3.94	0.102	0.004
2017	Census	0.204 (0.120)	8.46	0.097	0.138	0.210* (0.118)	8.46	0.082	0.088
2022	Census	0.148 (0.124)	13.09	0.239	0.451	0.173 (0.120)	13.09	0.157	0.293
2023	Annual	0.058 (0.139)	13.07	0.679	0.906	0.026 (0.137)	13.07	0.849	0.975

Notes: The baseline is the 1997 Census of Governments, the first full-Census wave in the source. Endpoints through 2022 are later full-Census waves; 2023 is shown only as a latest-year Annual Survey check. Treatment changes are also measured from 1997. The specification is the preferred joint historical-corridor long difference with direct terrain controls. Winsorization clips the outcome change at its sample-specific 1st and 99th percentiles. Stars follow the same weak-IV rule as Table G.2.