

Online Appendix for
The Impact of Data Centers on Local Economies

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. For house prices, the pooled outcome is $\log(HPI_{ct})$ and the long-difference outcome is $\log(HPI_{ct}) - \log(HPI_{c,1995})$.

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

$$Q_{ct} = \sum_i \mathbf{1}\{c(i) = c, 1995 \leq b_i \leq t\} r_i, \quad D_{ct} = \text{asinh}(Q_{ct}), \quad \Delta_t D_c = D_{ct} - D_{c,1995}. \quad (6)$$

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

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

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

$$A_c = \log(1 + N_c) - \log(1 + B_c). \quad (8)$$

We standardize the two lower-order components as

$$\tilde{C}_c = \frac{C_c - \bar{C}}{s_C}, \quad \tilde{A}_c = \frac{A_c - \bar{A}}{s_A}, \quad (9)$$

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

$$Z_c^{AN} = \tilde{C}_c \tilde{A}_c. \quad (10)$$

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

$$\tilde{Z}_c^{AN} = \frac{Z_c^{AN} - \overline{Z^{AN}}}{s_{Z^{AN}}}. \quad (11)$$

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 $\tilde{Z}_c^{AN}$ therefore states explicitly that the interaction itself is standardized.

The long-difference and balance regressions include $\tilde{C}_c$ and $\tilde{A}_c$ 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

$$K_e = d_e (1 + 400g_e^2), \quad (12)$$

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

$$Z_c^{LC} = \text{std}\{-\log(1 + d_c^{LC})\}. \quad (13)$$

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

$$H_c = \frac{\tilde{Z}_c^{AN} + Z_c^{LC}}{\sqrt{2}}. \quad (14)$$

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

$$s_c = \frac{H_c - \min_j H_j}{\max_j H_j - \min_j H_j}. \quad (15)$$

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,

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

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 $\tilde{Z}_c^{AN}$ 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, $\tilde{Z}_c^{AN}$ 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

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

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.

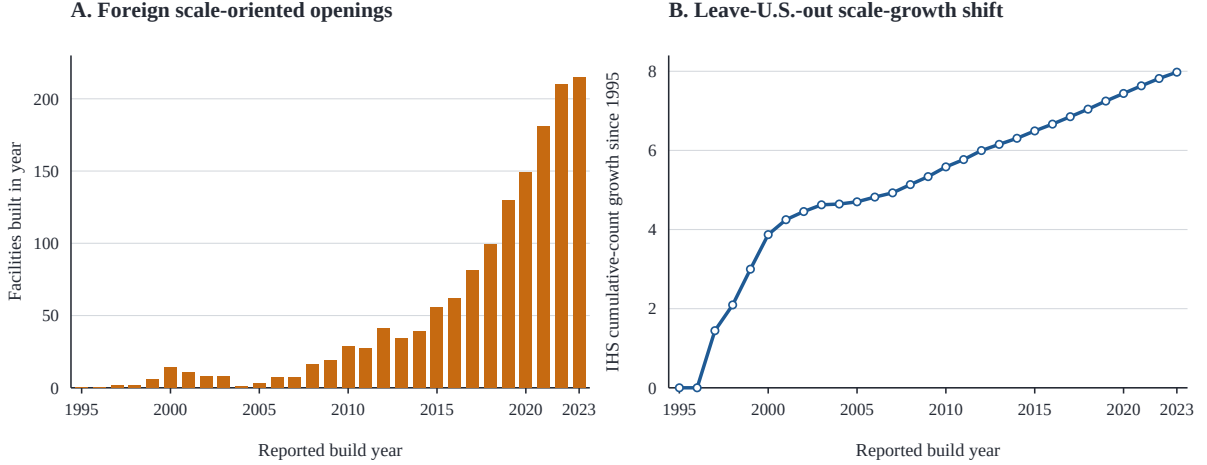


Figure A.1: Foreign growth in scale-oriented facilities

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

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

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

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.

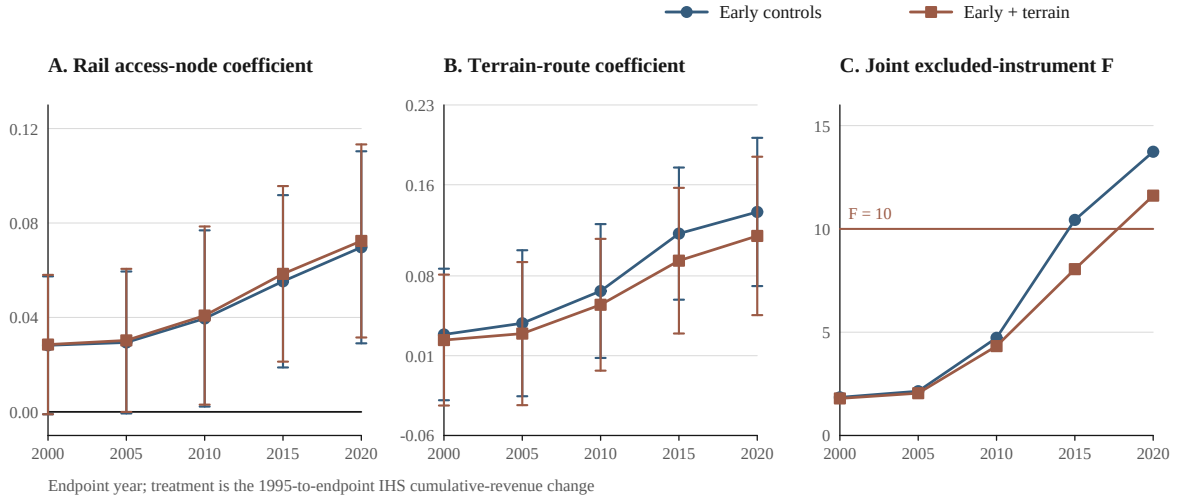


Figure B.1: The joint first stage strengthens with the cumulative horizon

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,

$$Z_{ct}^F = F_c G_t^{ROW}, \quad Z_{ct}^U = U_c G_t^{ROW}. \quad (19)$$

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
<i>Panel A: Full fiber/college-share support</i>						
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
<i>Panel B: Preferred-design common support</i>						
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 includes all counties with observed fiber proximity and the 1980 urban-college share. Panel B includes the subset of those counties with observed 1870 population and county boundaries stable by 1870, complete 1986 CBP and 1989 IRS controls, and matched ETOPO, 1870 railroad, and terrain-predicted-route geometry used by the preferred historical-corridor design. 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
	1986	0.065	(0.051)	18.25	0.212	0.392	1,547
Construction employment	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
	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
Establishments	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
	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
Annual payroll	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 elasticity	1986	-0.012	(0.007)	21.13	0.112	0.099	739
	1987	-0.007	(0.007)	21.13	0.287	0.232	739
	1988	0.001	(0.005)	21.13	0.917	0.630	739
	1989	0.007	(0.005)	21.13	0.216	0.480	739
	1990	0.008	(0.005)	21.13	0.126	0.396	739
	1991	0.007	(0.005)	21.13	0.141	0.423	739
Electricity price	1992	0.006	(0.004)	21.13	0.160	0.487	739
	1993	0.003	(0.003)	21.13	0.338	0.668	739
	1994	0.002	(0.002)	21.13	0.238	0.437	739
	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
Public-supply water withdrawals	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
	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. House prices use $\log HPI_{c,s} - \log HPI_{c,1995}$. 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	0.176	1,547
	1990	0.049	(0.026)	17.64	0.068	0.179	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
	1986	0.066	(0.051)	17.64	0.203	0.441	1,547
Construction employment	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
	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
Establishments	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
	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
Annual payroll	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 $GROW$ 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 elasticity	1986	-0.011	(0.007)	20.38	0.127	0.122	739
	1987	-0.007	(0.007)	20.38	0.309	0.279	739
	1988	0.000	(0.005)	20.38	0.938	0.614	739
	1989	0.007	(0.005)	20.38	0.225	0.466	739
	1990	0.008	(0.005)	20.38	0.133	0.349	739
	1991	0.007	(0.005)	20.38	0.157	0.361	739
Electricity price	1992	0.005	(0.004)	20.38	0.181	0.453	739
	1993	0.003	(0.003)	20.38	0.374	0.635	739
	1994	0.002	(0.002)	20.38	0.232	0.506	739
	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
Public-supply water withdrawals	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
	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. House prices use $\log HPL_{cs} - \log HPL_{c,1995}$. 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 two panels of Appendix Table E.3. 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 Panel A of Table E.3; 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 elasticity	2023	0.029***	(0.003)	< 0.001	0.9%	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 Panel B of Table E.3; cumulative data-center revenue is not instrumented. The implied-effect column evaluates each coefficient at the outcome sample's mean treatment change. House prices and 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	0.381	0.637	0.304**	0.203*
	(0.382)	(0.302)	(0.373)	(0.125)	(0.113)
	AR $p = 0.319$ $F = 1.85$	AR $p = 0.298$ $F = 2.14$	AR $p = 0.102$ $F = 4.73$	$p = 0.020$ $F = 10.44$	$p = 0.081$ $F = 13.74$
Construction employment	1.225	1.016	2.968*	1.110*	0.563
	(1.203)	(1.821)	(1.647)	(0.557)	(0.343)
	AR $p = 0.523$ $F = 1.85$	AR $p = 0.550$ $F = 2.14$	AR $p = 0.099$ $F = 4.73$	$p = 0.054$ $F = 10.44$	$p = 0.109$ $F = 13.74$
Establishments	0.056	0.193	0.224*	0.191*	0.187**
	(0.084)	(0.177)	(0.132)	(0.096)	(0.088)
	AR $p = 0.619$ $F = 1.85$	AR $p = 0.188$ $F = 2.14$	AR $p = 0.060$ $F = 4.73$	$p = 0.053$ $F = 10.44$	$p = 0.039$ $F = 13.74$
Annual payroll	0.805	0.703	0.757**	0.451**	0.342**
	(0.585)	(0.473)	(0.466)	(0.169)	(0.149)
	AR $p = 0.193$ $F = 1.85$	AR $p = 0.183$ $F = 2.14$	AR $p = 0.030$ $F = 4.73$	$p = 0.011$ $F = 10.44$	$p = 0.027$ $F = 13.74$
Tax returns	0.141***	0.140*	-0.139**	0.078	0.065
	(0.088)	(0.121)	(0.312)	(0.077)	(0.074)
	AR $p = 0.008$ $F = 1.85$	AR $p = 0.075$ $F = 2.14$	AR $p = 0.027$ $F = 4.70$	$p = 0.317$ $F = 10.44$	$p = 0.382$ $F = 13.74$
Adjusted gross income	0.347***	0.300***	-0.579	0.182**	0.142
	(0.200)	(0.198)	(0.698)	(0.085)	(0.093)
	AR $p = 0.003$ $F = 1.85$	AR $p = 0.004$ $F = 2.14$	AR $p = 0.586$ $F = 4.70$	$p = 0.040$ $F = 10.44$	$p = 0.135$ $F = 13.74$
Wages and salaries	0.355***	0.297**	-0.351**	0.179**	0.166*
	(0.196)	(0.171)	(0.660)	(0.085)	(0.092)
	AR $p = 0.001$ $F = 1.85$	AR $p = 0.019$ $F = 2.14$	AR $p = 0.015$ $F = 4.70$	$p = 0.042$ $F = 10.44$	$p = 0.080$ $F = 13.74$
House-price elasticity	0.139***	0.219**	0.086	0.129**	0.138***
	(0.099)	(0.137)	(0.060)	(0.054)	(0.051)
	AR $p = 0.003$ $F = 1.70$	AR $p = 0.010$ $F = 1.71$	AR $p = 0.181$ $F = 3.14$	AR $p = 0.015$ $F = 6.79$	AR $p = 0.000$ $F = 8.81$
Electricity-price elasticity	-0.009	-0.010	0.022	0.024	0.061
	(0.045)	(0.085)	(0.109)	(0.080)	(0.069)
	AR $p = 0.861$ $F = 2.19$	AR $p = 0.754$ $F = 2.47$	AR $p = 0.977$ $F = 4.91$	AR $p = 0.880$ $F = 9.92$	$p = 0.382$ $F = 11.01$
Public-supply water withdrawals	0.513	0.519	0.461	0.330*	0.306*
	(0.451)	(0.430)	(0.305)	(0.183)	(0.156)
	AR $p = 0.256$ $F = 1.85$	AR $p = 0.276$ $F = 2.14$	AR $p = 0.163$ $F = 4.73$	$p = 0.081$ $F = 10.44$	$p = 0.058$ $F = 13.74$

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; house prices and electricity prices are in natural logs, so each 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	0.416	0.729	0.344*	0.221*
	(0.382)	(0.324)	(0.417)	(0.149)	(0.121)
	AR $p = 0.417$ $F = 1.79$	AR $p = 0.308$ $F = 2.05$	AR $p = 0.116$ $F = 4.33$	AR $p = 0.060$ $F = 8.05$	$p = 0.076$ $F = 11.61$
Construction employment	0.967	0.925	3.110	1.273**	0.614*
	(1.134)	(1.875)	(1.789)	(0.565)	(0.345)
	AR $p = 0.600$ $F = 1.79$	AR $p = 0.430$ $F = 2.05$	AR $p = 0.106$ $F = 4.33$	AR $p = 0.036$ $F = 8.05$	$p = 0.083$ $F = 11.61$
Establishments	0.065	0.226	0.255*	0.223**	0.208**
	(0.083)	(0.184)	(0.141)	(0.102)	(0.091)
	AR $p = 0.654$ $F = 1.79$	AR $p = 0.206$ $F = 2.05$	AR $p = 0.065$ $F = 4.33$	AR $p = 0.027$ $F = 8.05$	$p = 0.028$ $F = 11.61$
Annual payroll	0.780	0.762	0.857**	0.514**	0.371**
	(0.589)	(0.511)	(0.487)	(0.199)	(0.166)
	AR $p = 0.266$ $F = 1.79$	AR $p = 0.198$ $F = 2.05$	AR $p = 0.041$ $F = 4.33$	AR $p = 0.029$ $F = 8.05$	$p = 0.031$ $F = 11.61$
Tax returns	0.159**	0.178*	-0.080**	0.115*	0.089
	(0.097)	(0.126)	(0.295)	(0.080)	(0.074)
	AR $p = 0.011$ $F = 1.79$	AR $p = 0.064$ $F = 2.05$	AR $p = 0.023$ $F = 4.28$	AR $p = 0.058$ $F = 8.05$	$p = 0.240$ $F = 11.61$
Adjusted gross income	0.369***	0.360***	-0.560	0.230***	0.177*
	(0.213)	(0.212)	(0.731)	(0.091)	(0.094)
	AR $p = 0.003$ $F = 1.79$	AR $p = 0.004$ $F = 2.05$	AR $p = 0.647$ $F = 4.28$	AR $p = 0.004$ $F = 8.05$	$p = 0.067$ $F = 11.61$
Wages and salaries	0.364***	0.334**	-0.265**	0.221***	0.195**
	(0.203)	(0.181)	(0.615)	(0.089)	(0.094)
	AR $p = 0.001$ $F = 1.79$	AR $p = 0.014$ $F = 2.05$	AR $p = 0.012$ $F = 4.28$	AR $p = 0.008$ $F = 8.05$	$p = 0.044$ $F = 11.61$
House-price elasticity	0.121***	0.201***	0.100**	0.135***	0.140***
	(0.093)	(0.131)	(0.070)	(0.060)	(0.054)
	AR $p = 0.001$ $F = 1.63$	AR $p = 0.006$ $F = 1.64$	AR $p = 0.025$ $F = 2.85$	AR $p = 0.003$ $F = 5.55$	AR $p = 0.000$ $F = 7.82$
Electricity-price elasticity	-0.010	-0.002	0.021	0.018	0.056
	(0.040)	(0.074)	(0.100)	(0.075)	(0.062)
	AR $p = 0.783$ $F = 2.13$	AR $p = 0.709$ $F = 2.37$	AR $p = 0.921$ $F = 4.57$	AR $p = 0.786$ $F = 7.76$	AR $p = 0.185$ $F = 9.10$
Public-supply water withdrawals	0.494	0.498	0.465	0.335	0.311*
	(0.451)	(0.434)	(0.320)	(0.198)	(0.163)
	AR $p = 0.253$ $F = 1.79$	AR $p = 0.280$ $F = 2.05$	AR $p = 0.160$ $F = 4.33$	AR $p = 0.140$ $F = 8.05$	$p = 0.065$ $F = 11.61$

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; house prices and electricity prices are in natural logs, so each 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.

Table E.3: Historical-corridor estimates: pooled and long differences

Panel A. Pooled annual estimates using scale-oriented foreign growth							
Outcome	End (1)	Estimate (2)	(SE) (3)	<i>p</i> -value (4)	AR <i>p</i> (5)	Implied effect at	
						Mean sample change (6)	Median increment at mean positive-revenue base (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 <i>F</i>				
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 [†]				
Panel B. 2SLS estimates at the latest available endpoint							
Outcome	End (1)	Estimate (2)	(SE) (3)	<i>p</i> -value (4)	AR <i>p</i> (5)	Implied effect at	
						Mean sample change (6)	Median increment at mean positive-revenue base (7)
Total employment	2023	0.274**	(0.122)	0.031	0.047	6.0%	2.6%
Construction employment	2023	0.600*	(0.314)	0.064	0.147	13.7%	5.9%
Establishments	2023	0.226**	(0.097)	0.025	0.028	5.0%	2.2%
Annual payroll	2023	0.360**	(0.145)	0.018	0.033	8.0%	3.5%
Tax returns	2022	0.108	(0.083)	0.200	0.086	2.3%	1.0%
Adjusted gross income	2022	0.231**	(0.113)	0.047	0.026	5.0%	2.2%
Wages and salaries	2022	0.190*	(0.098)	0.061	0.060	4.1%	1.8%
House-price elasticity	2023	0.142*** [†]	(0.058)	< 0.001	< 0.001	4.4%	1.4%
Electricity-price elasticity	2023	0.048 [†]	(0.063)	0.606	0.606	1.0%	0.5%
Public-supply withdrawals	2020	0.311*	(0.163)	0.065	0.089	6.7%	3.0%
Outcome sample	Start	End	First-stage <i>F</i>				
CBP outcomes	1995	2023	11.92				
IRS outcomes	1995	2022	11.87				
House prices	1995	2023	8.11 [†]				
Electricity prices	1995	2023	9.39 [†]				
Public-supply withdrawals	1995	2020	11.61				

Notes: Panel A reports pooled annual estimates. Annual panels begin in 1995; the public-supply panel begins in 2000. The instrumental variable is $s_c G_t$. Every specification includes county and state-by-year fixed effects. Outcomes use IHS except for logged house and electricity prices. First-stage statistics differ 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. Panel B reports a separate 1995-to-endpoint long-difference 2SLS regression in each row using the rail-coverage-by-access-node interaction and terrain-predicted route proximity jointly. All models use Early + terrain controls and state fixed effects. The endpoint is the final year available in the source. Treatment is IHS cumulative annualized county data-center revenue, measured in millions of dollars, in Panel A and its change from 1995 in Panel B. House and electricity prices enter in natural logs; Panel B uses their log changes from 1995, so at positive revenue values their coefficients are approximate elasticities. In both panels, a dagger marks an outcome sample with first-stage $F < 10$. For those rows, the p -value in column (4) of Panel B and the significance stars in both panels use the joint Anderson–Rubin test, which remains valid under weak identification. Column (4) of Panel A reports conventional p -values. Column (5) reports Anderson–Rubin p -values for every outcome in both panels. Standard errors are clustered by state. Column (6) of Panel A and column (6) of Panel B evaluate each coefficient at the unweighted mean sample change in IHS cumulative revenue among the outcome’s estimation counties. Column (7) of both panels uses $\Delta D = \text{asinh}(110) - \text{asinh}(100)$, a rounded \$10 million increment at a \$100 million base motivated by the median positive county-by-entry-year increment and mean positive-revenue base. We denote *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$.

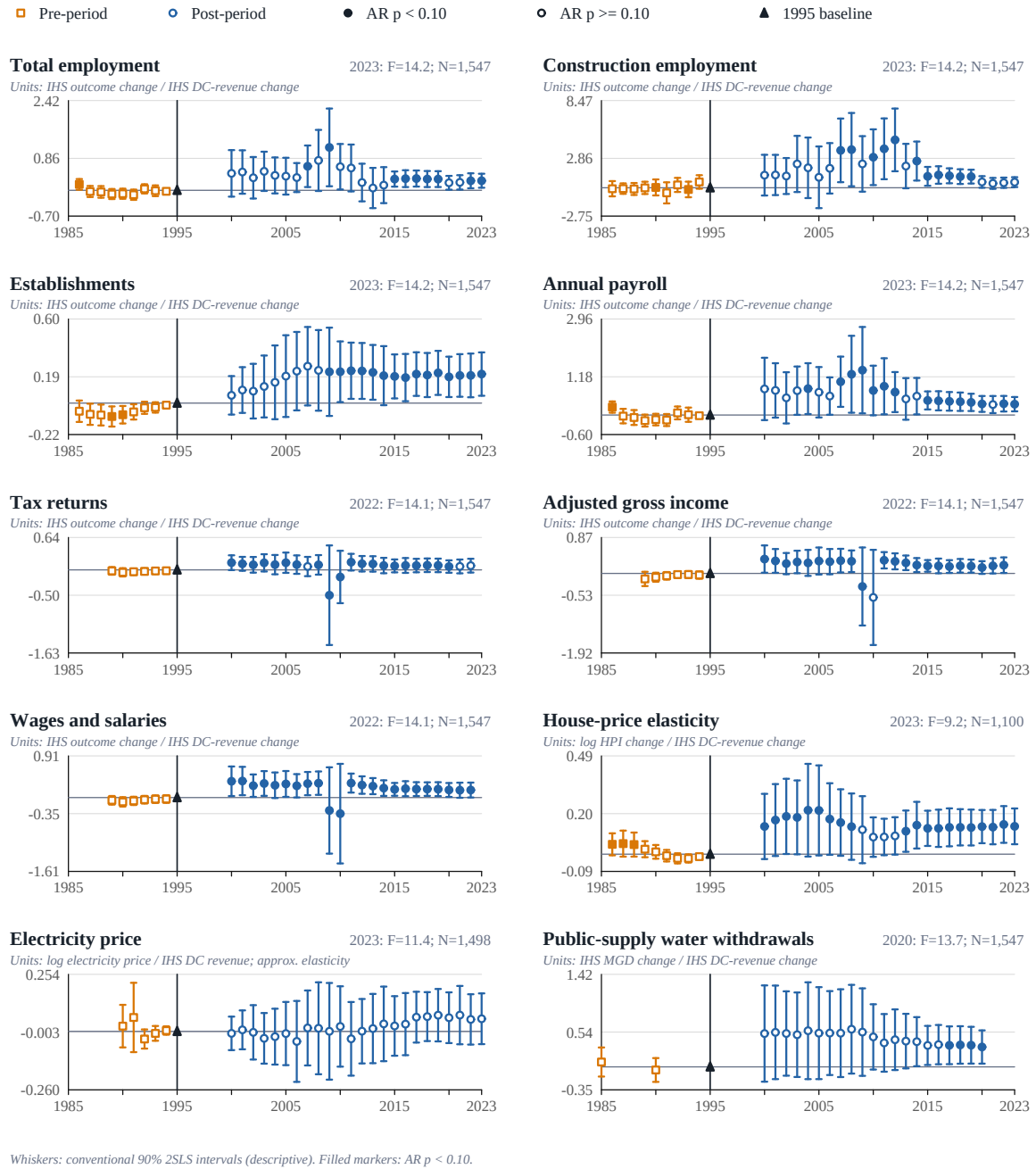


Figure E.1: Annual historical-corridor outcome profiles: Early controls

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. ●, ■: Anderson–Rubin $p < 0.10$; ○, □: Anderson–Rubin $p \geq 0.10$. For IHS outcomes the vertical axis has the same approximate elasticity interpretation at positive values; house and electricity prices use log differences and therefore have an approximate elasticity interpretation at positive revenue values.

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 Data centers, air pollution, and local economic activity

We examine whether data-center development affects fine particulate matter ($\text{PM}_{2.5}$) and ground-level ozone, and how these changes compare with local economic expansion. Table G.1 applies the paper’s pooled and preferred long-difference specifications in Panels A and B; Figure G.1 shows annual long differences and pre-period placebos with the preferred Early + terrain controls.

G.1 Data and specifications

Our main $\text{PM}_{2.5}$ measures begin in 1990, allowing the paper’s 1995 baseline and 1990–1994 placebos. The historical EPA CMAQ series extends through 2010; the combined historical/SatPM series extends through 2023.

The historical EPA series uses a consistent Community Multiscale Air Quality (CMAQ) simulation at 36-kilometer resolution, aggregated with Census 2000 population weights. Its published sample requires at least 20,000 residents in every year from 1990 through 2010; 1,190 counties overlap with the paper’s pooled sample.¹

The combined measure joins Meng et al.’s historical North American estimates for 1990–1999 with SatPM V6.NA.01 for 2000–2023, using the same fixed county-area weights for both 0.01-degree grids. The source changes in 2000. Pooled state-by-year fixed effects absorb changes common to counties within each state, but county-specific measurement differences can remain, and long differences span the two sources. We therefore treat this measure as suggestive evidence.²

Complementary CDC/EPA Downscaler data cover 2001–2022. We average daily population-weighted county predictions over complete calendar years: $\text{PM}_{2.5}$ in micrograms per cubic meter and daily maximum eight-hour ozone in parts per billion. Coverage includes all 1,577 pooled-sample counties through 2019 and 1,558 thereafter. Their long differences use 2001 as the baseline for both pollution and treatment, limiting comparability with the main estimates and providing no pre-1995 placebos.³

All pollution outcomes are logged. Treatment is the inverse hyperbolic sine of cumulative reported annualized data-center revenue, in millions of dollars. Pooled models use the historical exposure share interacted with scale-oriented foreign data-center growth, county fixed effects, and state-by-year fixed effects. Long differences use the rail-access interaction and terrain-predicted

¹Source: EPA ScienceHub, documented by Wyatt et al. (2020). The source’s mortality-data coverage rule determines the county sample, including the later-year population restriction. We use its pollution field, which measures county-scale ambient conditions.

²Sources: Meng et al. (2019), historical North American $\text{PM}_{2.5}$, and SatPM V6.NA.01, accessed September 17, 2026. Annual concentrations are in micrograms per cubic meter. We weight grid-cell intersections within fixed 2020 Census county boundaries, divide historical stored values by 100 as documented, and concatenate the series without transition calibration. Across 3,108 overlapping counties, the correlation of 2000–2016 log changes is 0.81; that of annual log changes is 0.74. Estimation retains the paper’s sample restrictions.

³Sources: CDC’s National Environmental Public Health Tracking Network, $\text{PM}_{2.5}$ and ozone, accessed September 16, 2026. Downscaler combines air-quality modeling and monitoring information; the downloadable metadata does not identify the population-weight vintage. Inference treats released predictions as observed outcomes, without incorporating prediction uncertainty.

route proximity as separate instruments, state fixed effects, and Early or Early + terrain controls.⁴ Inference is clustered by state.

G.2 Pooled estimates and long differences

The historical EPA CMAQ coefficient is 0.073 for 1995–2010 with Early + terrain controls (Anderson–Rubin (AR) $p = 0.020$). Its weak first stage ($F = 3.30$) and wide 95 percent AR set, $[0.006, 15.954]$, limit precision. The pooled estimate is 0.017 (AR $p = 0.222$).

For combined historical/SatPM PM_{2.5}, the pooled and preferred 1995–2023 coefficients are 0.149 and 0.167, with first-stage F statistics of 11.18 and 11.92 and AR p -values of 0.011 and 0.003. At the mean treatment change, they imply 3.3 and 3.6 percent higher concentrations. These positive estimates remain subject to the source-change and placebo qualifications. Table G.2, Panel A, reports the confidence sets.

The complementary CDC/EPA pooled estimates are also positive: 0.167 for PM_{2.5} and 0.106 for ozone (AR $p = 0.001$ and 0.011; $F = 9.07$), implying increases of 2.6 and 1.6 percent at the mean treatment change.⁵ Their preferred 2001–2022 long differences are less conclusive: PM_{2.5} does not reject zero (AR $p = 0.542$), while ozone’s 95 percent AR set is empty. An empty set means the joint instrument moments reject every constant treatment coefficient; rejection of zero therefore does not establish a positive effect.

G.3 Annual long differences and pre-period placebos

Figure G.1 presents the main historical measures in Panels A and B and complementary CDC/EPA measures in Panels C and D. Combined-series estimates have empty 95 percent AR sets in 2018 and 2020–2022, marked by diamonds; the 2023 set is nonempty. Earlier endpoints often have weak first stages and wide intervals.

Placebos use pollution changes from 1995 to each year in 1990–1994 and fixed 1995–2020 treatment (Table G.2, Panel B). The historical EPA 1994 placebo rejects zero under both control sets (AR $p = 0.022$ and 0.026); 1992 also rejects at 10 percent under Early controls. Combined-series 1991 and 1992 placebos reject under Early + terrain controls (AR $p = 0.031$ and 0.006; Early-control $p = 0.033$ and 0.062). These associations precede the source transition and caution against conclusive causal interpretation of the positive post-period estimates.

⁴Early controls include historical population, land area, water, ruggedness, temperature, the two lower-order rail components, 1986 employment, establishments and payroll, and 1989 tax returns, adjusted gross income and wages. Terrain controls add mean log elevation, elevation dispersion, and slope.

⁵The tables report unadjusted AR p -values. In the original fifteen-outcome pollution exercise, the Benjamini–Hochberg adjusted q -values for these pooled tests were 0.016 and 0.081.

Table G.1: PM_{2.5} and ozone: pooled estimates and latest long differences

Outcome / source	End	Estimate	(SE)	Conv. p	AR p	Arithmetic implication	
						Mean sample change	\$10m at \$100m base
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
<i>Panel A. Pooled annual estimates</i>							
Historical CMAQ PM _{2.5}	2010	0.017 [†]	(0.017)	0.304	0.222	0.3%	0.2%
Combined PM _{2.5}	2023	0.149**	(0.070)	0.041	0.011	3.3%	1.4%
CDC/EPA PM _{2.5}	2022	0.167*** [†]	(0.077)	0.036	0.001	2.6%	1.6%
CDC/EPA ozone	2022	0.106*** [†]	(0.052)	0.048	0.011	1.6%	1.0%
<i>Panel B. Latest long differences: Early + terrain controls</i>							
Historical CMAQ PM _{2.5}	2010	0.073*** [†]	(0.045)	0.113	0.020	1.1%	0.7%
Combined PM _{2.5}	2023	0.167***	(0.064)	0.013	0.003	3.6%	1.6%
CDC/EPA PM _{2.5}	2022	0.045 [†]	(0.083)	0.592	0.542	0.7%	0.4%
CDC/EPA ozone	2022	0.128*** ^{†‡}	(0.049)	0.013	0.004	2.0%	1.2%
<i>Samples and first stages (paired values: pooled / long difference)</i>							
Outcome / source	Period	First-stage F		Counties	Observations	States	
Historical CMAQ PM _{2.5}	1995–2010	7.84 / 3.30	1,190 / 1,163	19,040 / 1,163	38 / 38		
Combined PM _{2.5}	1995–2023	11.18 / 11.92	1,577 / 1,547	45,733 / 1,547	39 / 39		
CDC/EPA PM _{2.5}	2001–2022	9.07 / 8.47	1,577 / 1,528	34,637 / 1,528	39 / 39		
CDC/EPA ozone	2001–2022	9.07 / 8.47	1,577 / 1,528	34,637 / 1,528	39 / 39		

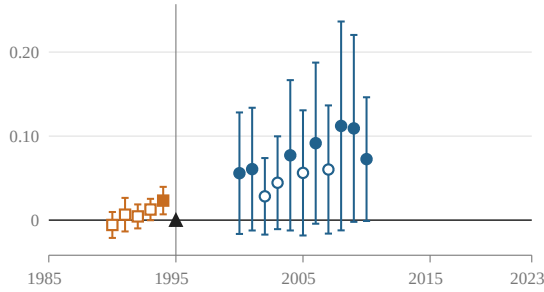
Notes: The first two sources support a 1995 baseline and pre-1995 placebos; CDC/EPA Downscaler estimates are complementary. Outcomes are log annual concentrations; treatment is the inverse hyperbolic sine (IHS) of cumulative reported annualized county data-center revenue in millions of dollars. Panel A includes county and state-by-year fixed effects and instruments treatment with historical exposure interacted with scale-oriented foreign data-center growth. Panel B differences outcomes and treatment over the stated period, includes Early + terrain controls and state fixed effects, and jointly uses the rail-access interaction and terrain-predicted route proximity as instruments. SEs and inference are clustered by state. Conventional p -values are two-sided; Anderson–Rubin (AR) p -values test a zero effect. Stars use AR inference: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$; † marks $F < 10$. Column (6) is $100\{\exp(\hat{\beta}\overline{\Delta D}) - 1\}$, using the unweighted mean start-to-end treatment change among unique sampled counties. Column (7) substitutes $\Delta D = \text{asinh}(110) - \text{asinh}(100) = 0.095306$, matching the main table’s \$10 million increment at a \$100 million base. Both are arithmetic point-estimate translations, with uncertainty assessed by AR inference. A double dagger marks an empty 95% AR set: the joint instrument moments reject every constant coefficient, so these magnitudes are illustrations, not identified effects; rejection of zero establishes no positive causal effect. Table G.2 reports the full sets. Combined PM_{2.5} uses historical North American data for 1990–1999 and SatPM V6.NA.01 for 2000–2023, aggregated over fixed county areas in original units without rescaling. This cross-source exercise is suggestive: geographically uneven measurement differences can remain after fixed effects.

□ Pre-period ○ Post-period ● AR $p < 0.10$
◆ Empty 95% AR set: the instruments reject every constant coefficient.

○ AR $p \geq 0.10$ ▲ Source baseline

A. PM_{2.5}: historical EPA CMAQ

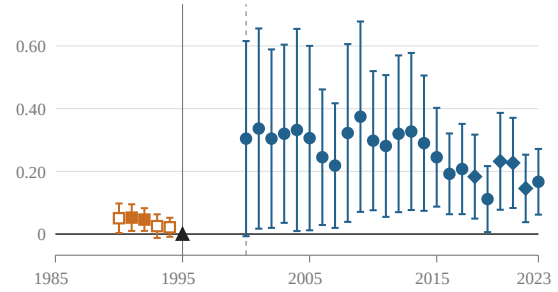
Baseline 1995; final 2010: $F=3.30$, $N=1,163$



1990-94 placebos; historical EPA series ends 2010.

B. PM_{2.5}: historical / SatPM

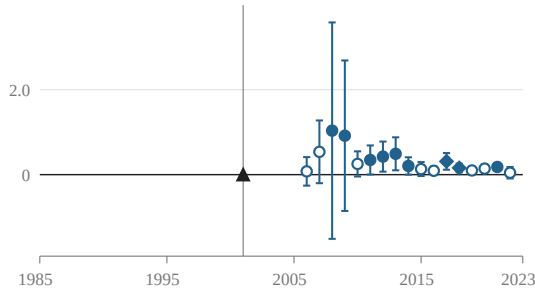
Baseline 1995; final 2023: $F=11.92$, $N=1,547$



1990-94 placebos; source changes in 2000 (dashed).

C. PM_{2.5}: CDC/EPA modeled

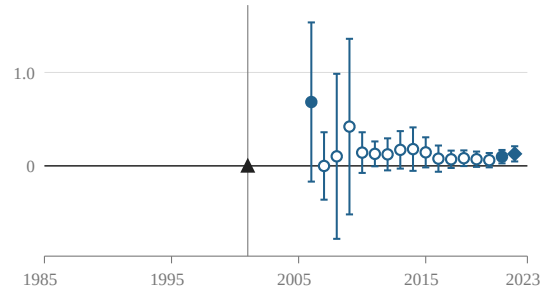
Baseline 2001; final 2022: $F=8.47$, $N=1,528$



Source begins 2001; no pre-1995 placebo.

D. Ozone: CDC/EPA modeled

Baseline 2001; final 2022: $F=8.47$, $N=1,528$



Source begins 2001; no pre-1995 placebo.

Vertical axis: change in log pollution / change in IHS data-center revenue. Bars: conventional 90% intervals.

Figure G.1: PM_{2.5} and ozone: annual long differences with Early + terrain controls

Notes: Blue markers show annual post-baseline endpoints, starting five years after each baseline; orange squares show pre-1995 placebos with fixed 1995–2020 treatment. Bars are conventional 90 percent 2SLS intervals; filled markers indicate AR $p < 0.10$ and diamonds identify empty 95 percent AR sets. Panel A uses historical EPA CMAQ (1995 baseline, ending 2010); Panel B uses combined historical/SatPM (1995 baseline, ending 2023), with the source change in 2000 marked by a dashed line. Panels C and D use complementary CDC/EPA measures (2001 baseline, no pre-1995 placebos). Baselines are normalizations. Regressions use logged pollution, both historical instruments, state fixed effects, and Early + terrain controls. Displayed F and N refer to the final endpoint. Placebo samples are fixed; post-period samples can vary by endpoint.

Table G.2: PM_{2.5} and ozone: AR confidence sets and pre-1995 placebos

<i>Panel A. 95% Anderson–Rubin confidence sets at the latest endpoint</i>			
Outcome / source	Pooled annual	Long difference: Early controls	Long difference: Early + terrain
Historical CMAQ PM _{2.5} 1995–2010	[−0.010, 0.104]	[0.007, 1.496]	[0.006, 15.954]
Combined PM _{2.5} 1995–2023	[0.035, 0.430]	[0.031, 0.388]	[0.069, 0.406]
CDC/EPA PM _{2.5} 2001–2022	[0.060, 0.578]	[−0.099, 0.238]	[−0.084, 0.243]
CDC/EPA ozone 2001–2022	[0.025, 0.353]	[0.085, 0.232]	∅ (empty)

<i>Panel B. Pre-1995 placebo estimates</i>				
Year	Early controls		Early + terrain controls	
	Estimate (SE)	AR p	Estimate (SE)	AR p
Historical CMAQ PM _{2.5} : 1,163 counties in 38 states $F = 11.19$ (Early), 9.08 (Early + terrain)				
1990	−0.008 (0.009)	0.324	−0.006 [†] (0.009)	0.429
1991	0.003 (0.011)	0.469	0.006 [†] (0.012)	0.638
1992	−0.004* (0.009)	0.055	0.004 [†] (0.009)	0.222
1993	0.012 (0.008)	0.280	0.013 [†] (0.008)	0.253
1994	0.024** (0.010)	0.022	0.023** [†] (0.010)	0.026
Combined PM _{2.5} : 1,547 counties in 39 states $F = 13.74$ (Early), 11.61 (Early + terrain)				
1990	0.047 (0.028)	0.200	0.050 (0.029)	0.112
1991	0.052** (0.025)	0.033	0.053** (0.026)	0.031
1992	0.044* (0.022)	0.062	0.046*** (0.022)	0.006
1993	0.022 (0.024)	0.503	0.025 (0.023)	0.466
1994	0.020 (0.018)	0.464	0.022 (0.018)	0.216

Notes: Panel A inverts the state-clustered Anderson–Rubin (AR) test at 5%; periods apply to outcomes and treatment. Specifications match Table G.1; Early controls omit the direct terrain block. Sets are pointwise and unadjusted for multiple testing. An empty set rejects every constant coefficient under the joint instrument moments and supplies no identified sign or magnitude. The CDC/EPA ozone Early-control 95% set is bounded, but its 90% set is empty. $(-\infty, +\infty)$ accepts every coefficient; a union of two intervals accepts both tails. Conventional coefficient/SE intervals are not weak-instrument robust. Panel B uses $\log P_{cs} - \log P_{c,1995}$ at year s , holding treatment fixed at its 1995–2020 IHS revenue change. Both historical instruments enter jointly with state fixed effects and the indicated controls. Within each source, counties and first-stage F are constant across displayed years; county samples are also constant across controls. SEs are state-clustered. Stars use unadjusted AR p -values: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$; † marks $F < 10$ and ‡ an empty 95% AR set. These are pre-expansion diagnostics, not effects of activity in the displayed years. Combined-series placebos and the 1995 baseline all use historical North American data; the post-2000 source switch remains a limitation for its long differences, as described in Table G.1. Both CDC/EPA series begin in 2001 and supply no pre-1995 placebos; unavailable observations are not filled with estimates or zeros.

G.4 Pollution relative to economic activity

We compare pollution and activity responses to data-center development, re-estimating both on exactly the same counties and years, with the same instruments and controls. For total employment, nominal annual payroll, or establishments, define

$$\eta_{P,A} = \frac{\beta_P}{\beta_A},$$

where β_P and β_A are the respective log-outcome coefficients. The ratio measures the percentage pollution response per percent activity response associated with the same data-center variation. Attributing pollution changes to activity alone would require an additional exclusion restriction.

The main combined-series pooled ratios are 0.40 for employment, 0.32 for payroll, and 0.58 for establishments; the corresponding 1995–2023 ratios are 0.85, 0.68, and 0.74. Complementary CDC/EPA pooled employment ratios are 0.43 for PM_{2.5} and 0.27 for ozone. The pooled point estimates imply less-than-proportional pollution increases, but all three employment ratios’ 95 percent AR sets include values above one. Table G.3 reports all coefficients, ratios, and confidence sets. Pooled ratio inference is weak-IV-robust AR; long-difference inference uses conventional Fieller sets, which are not weak-IV-robust and are unbounded for employment and payroll. The ratios retain the pollution estimates’ source, placebo, and empty-AR-set qualifications and do not establish a precise long-run activity–pollution elasticity.

Comparison with the literature. Antweiler, Copeland, and Taylor (2001) estimate an ambient SO₂–economic-scale elasticity of about 0.27, conditional on technique and composition. Their activity measure is GDP per square kilometer. Our pooled employment ratios of 0.40, 0.43, and 0.27 are of a similar order, although pollutants, activity measures, samples, and identifying variation differ.⁶

China’s experience illustrates how regulation and pollutant-specific changes can separate growth from pollution. Greenstone, He, Li, and Zou (2021) report a roughly 41 percent PM_{2.5} decline and 8.5 percent ozone increase during 2013–2018; these adjusted trends have no activity-growth denominator. Ebenstein et al. (2015) estimate how income and particulate pollution relate to life expectancy in China. For U.S. manufacturing, Shapiro and Walker (2018) document about 60 percent lower emissions during 1990–2008 despite rising real output, largely reflecting lower emissions intensity. These studies provide context rather than directly comparable elasticities: their outcomes are pollution trends, health, or source emissions, while ours are ambient concentrations relative to local activity gains.

⁶Their Table 1, Model B, p. 893, reports scale elasticities of 0.265 (random effects) and 0.266 (fixed effects). The outcome is log annual median SO₂; GDP per square kilometer combines city population density and national GDP per capita.

Table G.3: Pollution relative to the increase in economic activity

Pollutant	Activity	$\hat{\beta}_P$	$\hat{\beta}_A$	Ratio	95% confidence set	N
<i>Panel A. Pooled annual: AR inference for the ratio</i>						
Combined PM _{2.5} 1995–2023	Employment	0.149	0.370	0.402	[0.098, 1.277]	45,615
	Nominal payroll	0.149	0.469	0.318	[0.087, 0.743]	45,630
	Establishments	0.149	0.258	0.576	[0.157, 1.554]	45,733
PM _{2.5}	Employment	0.168	0.392	0.428	[0.158, 2.714]	34,537
	Nominal payroll	0.168	0.472	0.355	[0.135, 1.671]	34,552
	Establishments	0.167	0.313	0.534	[0.208, 1.434]	34,637
Ozone	Employment	0.106	0.392	0.270	[0.064, 1.706]	34,537
	Nominal payroll	0.106	0.472	0.224	[0.053, 1.074]	34,552
	Establishments	0.106	0.313	0.337	[0.097, 0.785]	34,637
<i>Panel B. Long differences: conventional Fieller inference</i>						
Combined PM _{2.5} 1995–2023	Employment	0.165	0.196	0.845	$(-\infty, -3.663]$ $\cup [0.216, +\infty)$ $(-\infty, -10.821]$	1,542
	Nominal payroll	0.165	0.245	0.675	$\cup [0.175, +\infty)$	1,542
	Establishments	0.167	0.226	0.737	[0.238, 3.770]	1,547
PM _{2.5}	Employment	0.044	0.171	0.260	$(-\infty, +\infty)$	1,526
	Nominal payroll	0.044	0.204	0.218	$(-\infty, +\infty)$	1,526
	Establishments	0.045	0.275	0.162	$[-0.562, 46.869]$	1,528
Ozone	Employment	0.128 [‡]	0.171	0.750 [‡]	$(-\infty, -0.620]$ $\cup [0.205, +\infty)$ $(-\infty, -0.704]$	1,526
	Nominal payroll	0.128 [‡]	0.204	0.629 [‡]	$\cup [0.171, +\infty)$	1,526
	Establishments	0.128 [‡]	0.275	0.467 [‡]	[0.181, 36.300]	1,528

Notes: The first block in each panel uses the main combined historical/SatPM PM_{2.5} series for 1995–2023. The following blocks provide complementary CDC/EPA evidence for 2001–2022, using the natural log of population-weighted annual PM_{2.5} ($\mu\text{g}/\text{m}^3$) or ozone (ppb). Activity is the natural log of total employment, nominal annual payroll, or establishments. Each row re-estimates both outcome responses to data-center development on exactly the same sample. Columns $\hat{\beta}_P$ and $\hat{\beta}_A$ are those two 2SLS coefficients; their ratio measures the percent pollution response per percent activity response associated with data-center development. It is not a separately identified causal effect of activity on pollution. Payroll is measured in current dollars.

Panel A includes county and state-by-year fixed effects and uses the paper’s single annual rail-access shift-share instrument. Treatment is IHS cumulative reported annualized county data-center revenue. Panel B uses changes in both outcomes and treatment over the indicated source period, Early + terrain controls, state fixed effects, and the two long-difference instruments. Inference is clustered by state. N denotes county-years in Panel A and counties in Panel B. In the CDC/EPA rows, employment/payroll lose 100/85 observations from the pooled pollution sample and two counties each from the long-difference sample; establishments use the full corresponding samples.

Panel A reports weak-instrument-robust Anderson–Rubin (AR) confidence sets for the single-instrument response ratio. Panel B reports conventional joint-normal Fieller sets, conditional on the estimated 2SLS responses and their joint state-clustered covariance; these are not weak-instrument-robust. Unbounded and disjoint sets retain all admitted values. A double dagger marks an empty 95% AR set for the underlying data-center-to-pollution coefficient on that row’s common sample: the joint instrument moments reject every constant coefficient. Ratios carrying this marker are arithmetic illustrations; normalization does not resolve that identification warning.

H 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 [H.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 H.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 $100[\exp(\hat{\beta} \Delta \bar{D}) - 1]$ 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 [H.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 H.2: Property-tax revenue: selected 1995-to-endpoint long differences

End	Release	Early controls, IHS			Preferred controls, IHS			Preferred, log positive		
		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 H.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 H.3: Sensitivity: 1997 full-Census baseline

End	Release	IHS change				Winsorized IHS change			
		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 H.2.

I Not Just Colocation: Data-Center Driven Growth

This section examines whether the historical-rail design could be capturing broader digital-sector development or modern interconnection hubs instead of data-center-driven growth. We report three complementary checks to further support the role of data-center investment in the local economic expansion documented above. First, before a county’s first mapped data center, we find no statistically significant relationship between the annual instrument and activity in other digital industries. This holds across all 21 reduced-form tests covering employment, establishments, and payroll in six non-hosting digital sectors, including telecommunications, software publishing, and computer-systems design, and their composite. The timing makes this evidence particularly informative: as Appendix Figure I.1 shows, a substantial share of counties first host a mapped data center well after 2010, long after the initial internet boom. For these counties, the pre-entry window therefore includes years of internet-era economic activity. This placebo period helps distinguish the local gains associated with data-center development from the earlier expansion of the internet economy, reinforcing the interpretation of data-center investment as a distinct driver of local growth.

Second, we use PeeringDB, a public directory of internet infrastructure, to identify interconnection facilities: physical sites where, for example, telecommunications carriers and internet service providers connect their networks and exchange traffic. We exclude counties containing listed facilities that cannot be matched to our data-center records under conservative matching rules. The resulting pooled point estimates are *larger* for every outcome and most point estimates also increase in the preferred long-difference design. Finally, all ten outcome estimates remain positive in both designs after flexibly controlling for the size and composition of the county’s 1995 digital economy.

I.1 Before entry: Non-hosting digital growth

To look for evidence of a broader digital-growth channel that predates the first mapped local data-center build, we examine non-hosting digital outcomes in the pre-entry sample. For counties that eventually host a facility in the DCKB, we retain only years strictly before the first mapped build; counties that never host a mapped facility remain in the sample throughout 1998–2023. Figure I.1 summarizes the timing of first-entry across counties.

The outcomes are inverse-hyperbolic-sine (IHS) employment, establishments, and annual payroll in six non-hosting digital sectors—telecommunications, software publishing, computer-systems design, internet publishing and media, call centers, and securities markets—and in a broad non-hosting digital composite. Data processing and hosting are excluded from these outcomes. The regressions use the paper’s annual instrument, county fixed effects, state-by-year fixed effects, and state-clustered inference.

Table I.1 reports the complete 21-outcome reduced-form screen. None of the 21 raw p -values is below 0.10; the smallest is 0.227. After adjusting jointly for the 21 tests using the Benjamini–Hochberg false-discovery-rate procedure, the smallest adjusted q -value is 0.812. In a separate

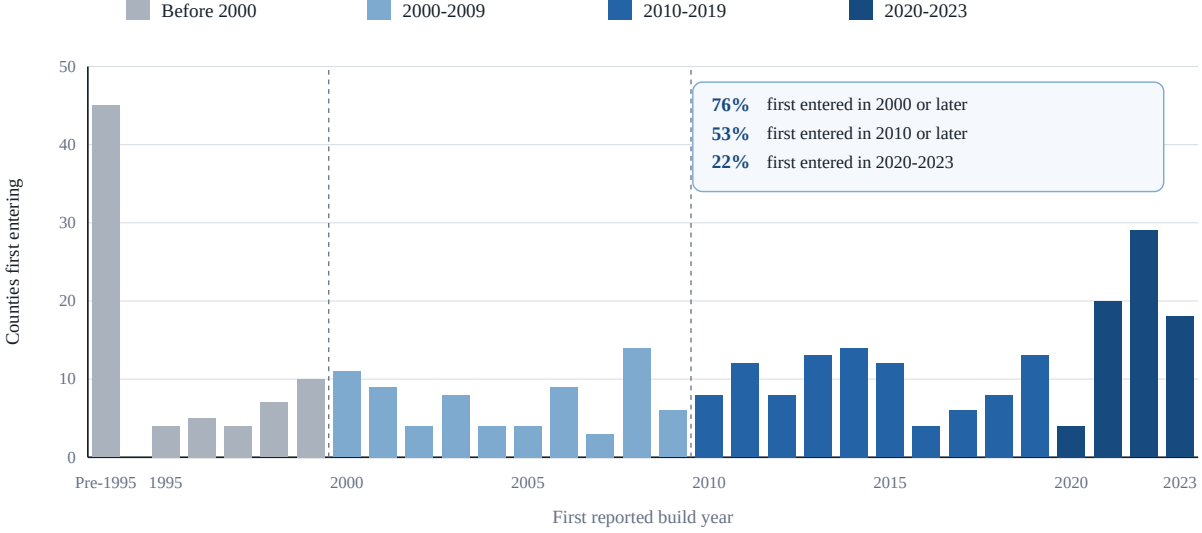


Figure I.1: Timing of first mapped data-center entry across counties

Notes: Each county is assigned the minimum reported build year among its mapped DCKB facilities. The figure includes 316 U.S. counties whose first mapped build is no later than 2023; earlier years are grouped. Of these counties, 76.3 percent first enter in 2000 or later, 53.5 percent in 2010 or later, and 22.5 percent between 2020 and 2023. Within the paper’s preferred historical-share support, the corresponding shares are 73.6 and 52.1 percent, and the median first-entry year is 2010. Source: 451 Research Datacenter Knowledge Base.

never-host-only sample, none of the 21 raw p -values is below 0.10 and the smallest adjusted q -value is 0.904.

Independent screens point in the same direction. The official November 1994 TOP500 screen uses county presence, site and system counts, and computing performance for institutional and research supercomputers to ask whether the historical measures simply identify pre-existing centers of advanced computing; its minimum adjusted q -values are 0.971 with Early controls and 0.978 with Early + terrain controls. The project FCC business-fiber screen measures county availability shares at 100/20 and 1000/100 Mbps, both in the full sample and after excluding every DCKB-host county; its joint p -values range from 0.359 to 0.870. The Census screen uses five household adoption shares from ACS table B28002—Internet subscription, broadband of any type, a cellular data plan, a cellular-only plan, and cable, fiber-optic, or DSL service—and retains only county-years before the first mapped DCKB build; its p -values range from 0.229 to 0.673.

Table I.1: Pre-entry reduced forms for non-hosting digital activity

Sector or composite	Employment	Establishments	Annual payroll
Telecommunications	0.147 (0.130) $p = .266, q = .812$	0.058 (0.066) $p = .387, q = .812$	0.283 (0.261) $p = .285, q = .812$
Software publishers	0.003 (0.054) $p = .952, q = .952$	-0.003 (0.032) $p = .925, q = .952$	0.021 (0.102) $p = .838, q = .952$
Computer-systems design	0.090 (0.119) $p = .458, q = .812$	0.004 (0.066) $p = .950, q = .952$	0.182 (0.218) $p = .410, q = .812$
Internet publishing/media	0.034 (0.059) $p = .562, q = .812$	-0.021 (0.030) $p = .475, q = .812$	0.082 (0.117) $p = .487, q = .812$
Call centers	0.094 (0.078) $p = .234, q = .812$	-0.018 (0.033) $p = .580, q = .812$	0.145 (0.118) $p = .227, q = .812$
Securities markets	-0.061 (0.096) $p = .528, q = .812$	-0.064 (0.062) $p = .308, q = .812$	-0.064 (0.201) $p = .751, q = .927$
Broad non-hosting digital composite	0.109 (0.137) $p = .431, q = .812$	0.026 (0.057) $p = .650, q = .853$	0.147 (0.260) $p = .576, q = .812$

Notes: Each cell reports the reduced-form coefficient, its state-clustered standard error in parentheses, the raw p -value, and the Benjamini–Hochberg q -value across the 21 tests. The q -value is the corresponding false-discovery-rate-adjusted p -value. The coefficient is the relationship between the paper’s annual instrument and the listed IHS outcome before the county’s first mapped data center; it is not a second-stage treatment-effect estimate. The common sample contains 38,648 county-year observations, 1,539 counties, and 39 state clusters from 1998 through 2023. All regressions include county and state-by-year fixed effects.

I.2 Holding the early digital economy fixed

This exercise asks whether the estimated effects reflect differences in the size or composition of counties' digital economies at the beginning of the treatment period. We construct 24 IHS-transformed 1995 baseline-at-onset controls: employment, establishments, and annual payroll in each of eight categories. Because 1995 is the first year of the treatment window, these are baseline-at-onset rather than strictly pre-treatment measures. The categories are telecommunications; computer services excluding data processing; software publishing; computer-systems design; information retrieval; a call-center and business-services proxy; securities brokerage; and the broad non-hosting digital composite. The closest historical data-processing and hosting analogue is removed from the relevant aggregates so that the controls absorb competing digital specialization without directly controlling for the activity being studied.

In the pooled annual regressions, each 1995 characteristic is interacted with year indicators, allowing counties with different initial digital profiles to follow different outcome paths. In the long-difference regressions, the 24 starting levels enter directly, together with two historical establishment changes ending in 1995: information retrieval from 1988 to 1995 and the non-hosting composite from 1986 to 1995. The long-difference specifications retain the preferred Early + terrain controls. All added controls are complete, so the county samples do not shrink.

Table [I.2](#) reports the baseline and saturated-control results. All ten pooled and all ten long-difference estimates remain positive. Using weak-IV-robust Anderson–Rubin inference, the zero-effect null is rejected at the 10 percent level for seven of ten pooled outcomes and six of ten long-difference outcomes after adding the controls. The median first-stage statistic falls to 2.70 in the pooled specifications and 3.19 in the long differences. Thus, this exercise is reassuring about the broad direction of the results and the Anderson–Rubin evidence, but it does not support treating movements in the exact point estimates as precise.

Table I.2: Outcome estimates after absorbing the early digital economy

Outcome (long-difference endpoint)	Pooled baseline	Pooled + digital controls	Long difference baseline	Long difference + digital controls
Total employment (2023)	0.430 $F = 11.18; p_{AR} = .005$	1.080 $F = 2.78; p_{AR} = .003$	0.274 $F = 11.92; p_{AR} = .047$	0.321 $F = 3.19; p_{AR} = .157$
Construction employment (2023)	1.389 $F = 11.18; p_{AR} = .013$	1.819 $F = 2.78; p_{AR} = .119$	0.600 $F = 11.92; p_{AR} = .147$	0.768 $F = 3.19; p_{AR} = .272$
Establishments (2023)	0.258 $F = 11.18; p_{AR} = .003$	0.558 $F = 2.78; p_{AR} = .006$	0.226 $F = 11.92; p_{AR} = .028$	0.348 $F = 3.19; p_{AR} = .028$
Annual payroll (2023)	0.576 $F = 11.18; p_{AR} < .001$	1.559 $F = 2.78; p_{AR} < .001$	0.360 $F = 11.92; p_{AR} = .033$	0.393 $F = 3.19; p_{AR} = .142$
Tax returns (2022)	0.193 $F = 10.80; p_{AR} = .016$	0.332 $F = 2.62; p_{AR} = .046$	0.108 $F = 11.87; p_{AR} = .086$	0.216 $F = 3.19; p_{AR} = .028$
Adjusted gross income (2022)	0.210 $F = 10.80; p_{AR} = .028$	0.426 $F = 2.62; p_{AR} = .083$	0.231 $F = 11.87; p_{AR} = .026$	0.372 $F = 3.19; p_{AR} = .007$
Wages and salaries (2022)	0.211 $F = 10.80; p_{AR} = .008$	0.431 $F = 2.62; p_{AR} = .034$	0.190 $F = 11.87; p_{AR} = .060$	0.310 $F = 3.19; p_{AR} = .026$
House prices (2023)	0.037 $F = 11.23; p_{AR} = .478$	0.078 $F = 2.41; p_{AR} = .553$	0.142 $F = 8.11; p_{AR} < .001$	0.166 $F = 2.61; p_{AR} = .004$
Electricity-price elasticity (2023)	0.042 $F = 8.79; p_{AR} = .550$	0.113 $F = 3.13; p_{AR} = .520$	0.048 $F = 9.39; p_{AR} = .606$	0.078 $F = 2.55; p_{AR} = .555$
Public-supply water withdrawals (2020)	0.247 $F = 6.79; p_{AR} = .009$	0.788 $F = 0.79; p_{AR} = .005$	0.311 $F = 11.61; p_{AR} = .089$	0.555 $F = 3.08; p_{AR} = .070$

Notes: Each cell reports the 2SLS coefficient, the excluded-instrument first-stage statistic F , and the weak-IV-robust Anderson–Rubin p -value for a zero treatment effect. The added-control columns absorb the 24 IHS-transformed 1995 baseline-at-onset digital-sector levels described in the text. In the pooled regressions, those levels are interacted with year indicators. The long-difference regressions also absorb the two pre-1995 establishment changes described in the text and retain the paper’s preferred Early + terrain controls. House prices enter in natural logs; the long-difference columns use $\log HPI_{c,2023} - \log HPI_{c,1995}$. Inference is clustered by state. The added controls are complete and therefore do not change the corresponding baseline samples.

I.3 Removing likely modern interconnection hubs

This exercise tests whether the results are driven by a small set of modern interconnection facilities, backbone nodes, or data centers missing from the DCKB treatment data. We first use a 2026 PeeringDB snapshot and compare every listed U.S. facility with every U.S. DCKB record. A PeeringDB facility is treated as matched when it lies within five kilometers of a DCKB record, has the same normalized address, or has a sufficiently similar name in the same locality. Of the 146 facilities that fail all three rules, 132 can be mapped defensibly to 117 counties. We drop every county in the corresponding estimation sample that contains at least one of these unmatched facilities. This removes 65 of 1,577 counties in the pooled estimation sample and 62 of 1,547 counties in the complete long-difference estimation sample; outcome-specific samples vary. The outcomes and controls are otherwise unchanged, and the long differences retain the preferred Early + terrain controls.

Table I.3 reports the targeted PeeringDB exclusion. All ten coefficient signs remain positive in both designs, and the Anderson–Rubin test rejects a zero effect at the 10 percent level for eight of ten outcomes in each. First-stage statistics range from 3.50 to 7.07 in the pooled regressions and from 5.34 to 9.35 in the long differences.

The underlying modern-facility diagnostic is informative but deliberately mixed. In the full sample, the two historical components jointly predict 14 of the 16 present-day PeeringDB outcomes at an adjusted q -value below 0.10 under each control set. The 16 outcomes comprise counts and presence indicators for PeeringDB facilities, IXPs, hosted networks, PeeringDB Content networks used as a CDN/edge proxy, reported carriers, and facilities conservatively unmatched to DCKB. The relationship is concentrated in DCKB-host counties: the two outcomes limited to conservatively unmatched facilities have a minimum adjusted q -value of 0.202, and after all DCKB-host counties are removed, none of the 16 PeeringDB outcomes has $q < 0.10$; the minimum is 0.269. This pattern is more consistent with DCKB colocation or downstream interconnection investment than with a clearly separate omitted digital sector, although the 2026 snapshot cannot by itself establish timing.

Table I.3: Outcome estimates after excluding counties with unmatched PeeringDB facilities

Outcome (long-difference endpoint)	Pooled baseline	Pooled: exclude PeeringDB	Long difference baseline	Long difference: exclude PeeringDB
Total employment (2023)	0.430 $F = 11.18; p_{AR} = .005$	0.691 (1.61 $\times$) $F = 3.74; p_{AR} = .004$	0.274 $F = 11.92; p_{AR} = .047$	0.308 (1.13 $\times$) $F = 6.63; p_{AR} = .054$
Construction employment (2023)	1.389 $F = 11.18; p_{AR} = .013$	1.946 (1.40 $\times$) $F = 3.74; p_{AR} = .021$	0.600 $F = 11.92; p_{AR} = .147$	0.753 (1.26 $\times$) $F = 6.63; p_{AR} = .120$
Establishments (2023)	0.258 $F = 11.18; p_{AR} = .003$	0.399 (1.54 $\times$) $F = 3.74; p_{AR} = .004$	0.226 $F = 11.92; p_{AR} = .028$	0.251 (1.11 $\times$) $F = 6.63; p_{AR} = .036$
Annual payroll (2023)	0.576 $F = 11.18; p_{AR} < .001$	0.920 (1.60 $\times$) $F = 3.74; p_{AR} < .001$	0.360 $F = 11.92; p_{AR} = .033$	0.384 (1.07 $\times$) $F = 6.63; p_{AR} = .043$
Tax returns (2022)	0.193 $F = 10.80; p_{AR} = .016$	0.307 (1.59 $\times$) $F = 3.53; p_{AR} = .017$	0.108 $F = 11.87; p_{AR} = .086$	0.094 (0.87 $\times$) $F = 6.66; p_{AR} = .056$
Adjusted gross income (2022)	0.210 $F = 10.80; p_{AR} = .028$	0.333 (1.58 $\times$) $F = 3.53; p_{AR} = .036$	0.231 $F = 11.87; p_{AR} = .026$	0.228 (0.99 $\times$) $F = 6.66; p_{AR} = .026$
Wages and salaries (2022)	0.211 $F = 10.80; p_{AR} = .008$	0.330 (1.56 $\times$) $F = 3.53; p_{AR} = .013$	0.190 $F = 11.87; p_{AR} = .060$	0.185 (0.97 $\times$) $F = 6.66; p_{AR} = .072$
House prices (2023)	0.037 $F = 11.23; p_{AR} = .478$	0.060 (1.62 $\times$) $F = 3.50; p_{AR} = .452$	0.142 $F = 8.11; p_{AR} < .001$	0.186 (1.31 $\times$) $F = 5.34; p_{AR} < .001$
Electricity-price elasticity (2023)	0.042 $F = 8.79; p_{AR} = .550$	0.058 (1.39 $\times$) $F = 7.07; p_{AR} = .592$	0.048 $F = 9.39; p_{AR} = .606$	0.059 (1.22 $\times$) $F = 9.35; p_{AR} = .444$
Public-supply water withdrawals (2020)	0.247 $F = 6.79; p_{AR} = .009$	0.310 (1.25 $\times$) $F = 4.73; p_{AR} = .010$	0.311 $F = 11.61; p_{AR} = .089$	0.388 (1.25 $\times$) $F = 6.54; p_{AR} = .078$

Notes: Each cell reports the 2SLS coefficient, the excluded-instrument first-stage statistic F , and the weak-IV-robust Anderson–Rubin p -value for a zero treatment effect. Parenthetical multiples in the exclusion columns are ratios of the exclusion-sample point estimate to the corresponding baseline point estimate; they do not measure precision. The exclusion removes counties in the corresponding estimation sample that contain a PeeringDB facility that cannot be matched to a DCKB record under any of the three rules described in the text. House prices enter in natural logs; the long-difference columns use $\log HPI_{c,2023} - \log HPI_{c,1995}$. Inference is clustered by state.

The targeted PeeringDB exercise removes possible unmeasured facilities while preserving substantially more identifying variation than broad geographic exclusions. For completeness, Table I.4 reports progressively wider stress tests based on the published 2015 InterTubes backbone map. The map contains 273 nodes. Under the project’s existing pre-2000 verification rule, a node is excluded when the supporting project documentation provides no clear evidence of an associated long-haul tube or conduit before 2000. The rule removes six nodes, leaving 267 retained nodes in 232 counties. The 25- and 50-kilometer rules use distances between each county centroid and the centroid of the nearest retained-node county, with retained-node counties assigned a distance of zero.

Table I.4: Modern-hub exclusion ledger and loss of identifying support

Exclusion	Counties removed pooled/LD	Positive signs pooled/LD	First-stage F range pooled/LD	Interpretation
Unmatched PeeringDB facilities	65 / 62	10/10 / 10/10	3.50–7.07 / 5.34–9.35	Best targeted check
InterTubes node counties	108 / 103	10/10 / 10/10	1.50–3.95 / 1.14–3.64	Signs persist; F is weak
Within 25 km	150 / 145	10/10 / 10/10	2.43–4.10 / 2.03–2.25	Signs persist; F is weak
Within 50 km	441 / 426	0/10 / 3/10	0.19–0.32 / 0.22–0.49	Not interpretable: relevance collapses

Notes: The sign counts and first-stage ranges are computed across the paper’s ten outcomes. The paired entries report pooled and preferred long-difference results, respectively. The radius rows exclude counties whose centroid lies within the stated distance of the centroid of any retained-node county; retained-node counties have distance zero. The 25-kilometer samples retain 38 state clusters and the 50-kilometer samples retain 36. At 50 kilometers, the instrument has essentially stopped predicting data-center development, so those coefficients do not provide an interpretable alternative causal estimate. Because the hub measures are modern snapshots and the exclusions change the set of counties represented, these exercises are stress tests rather than exogenous validation.

J Spatial equilibrium

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. Workers choose 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 interpret locations in the model as counties in the data.

We assume locations are heterogeneous in their suitability to install a data center. We sketch here two possible comparative static exercises that in such a model rationalize the entry of a data center in a location. The first one is a technological development in the demand for 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 the prices of other inputs. Higher wages attract workers from other locations, and higher population levels and incomes increase 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, 2010](#)). 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 first pass, 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 entry 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 technology-driven explanation will be welfare improving – or more precisely it will attract workers to the locations until it dissipates the marginal welfare effects, and produce gains for 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 is 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 congestion effects.

The West Des Moines case suggests that a mix of the two explanations may be at play. It also suggests a more nuanced view of the local subsidies 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, in the absence of externalities, will be also negative, i.e. will not compensate the loss of 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 patterns are consistent with the qualitative patterns 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. This framework rationalizes the joint outcome pattern, while the magnitudes of the responses suggest that fiscal-infrastructure spillovers or other local multipliers may provide substantial amplification beyond the direct-demand channel. We leave the analysis of these considerations for future research.

K 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 K.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.

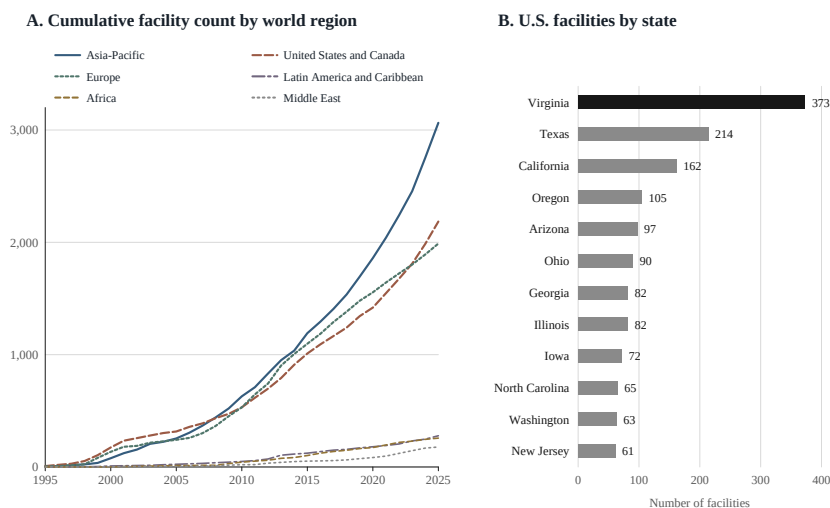


Figure K.1: Data-center growth and geographic concentration

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.

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 K.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

⁷To 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.

the U.S. is over 20 billion dollars and operational capacity surpasses 90 million square feet.

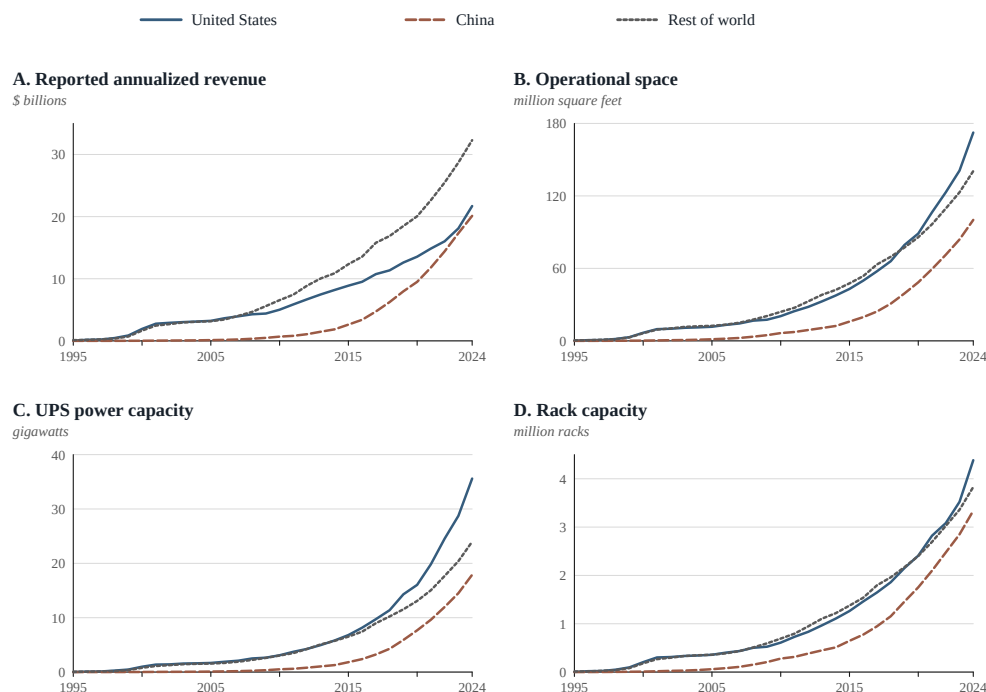


Figure K.2: Reported scale of data-center facilities built by year

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.

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 K.3 for the U.S. These scale-oriented facilities are designed to accommodate large and expandable computing loads, with 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.

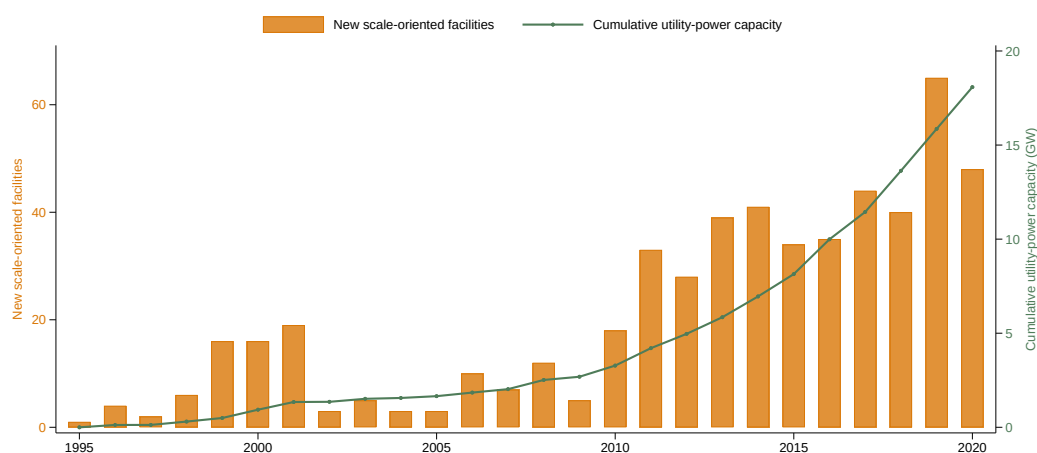


Figure K.3: Expansion of scale-oriented data centers in the United States

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.

References

- Bartik, T.J., 1991. Who Benefits from State and Local Economic Development Policies? W.E. Upjohn Institute for Employment Research, Kalamazoo, MI.
- 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.
- Kline, P., Moretti, E., 2014. People, places, and public policy: Some simple welfare economics of local economic development programs. *Annual Review of Economics* 6, 629–662. doi:[10.1146/annurev-economics-080213-041024](https://doi.org/10.1146/annurev-economics-080213-041024).
- Moretti, E., 2010. Local multipliers. *American Economic Review* 100, 373–377. doi:[10.1257/aer.100.2.373](https://doi.org/10.1257/aer.100.2.373).
- Roback, J., 1982. Wages, rents, and the quality of life. *Journal of Political Economy* 90, 1257–1278. doi:[10.1086/261120](https://doi.org/10.1086/261120).
- 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](https://doi.org/10.1257/aer.20141702).