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

Fernando Alvarez, University of Chicago
David Argente, Yale University
Joyce Chow, Yale University
Diana Van Patten, Yale University

May 2026

Abstract

Data centers are the physical infrastructure behind cloud computing, artificial intelligence, and enterprise software. The rapid diffusion of artificial intelligence (AI) is intensifying demand for compute, accelerating investment in data centers, and raising concerns about the local economic and environmental footprint of these facilities. Their expansion creates a local policy tradeoff. A data center can bring capital investment, construction activity, and specialized employment, but it can also increase demand for electricity, land, and grid capacity. This paper studies these effects at the U.S. county level. We assemble a facility-level panel of global data centers with precise coordinates, scale metrics, and annualized revenue. We map facilities to U.S. counties and combine them with County Business Patterns, county-level IRS income, county-level house prices, and electricity prices. To address endogenous siting, we instrument for data center growth using two shift-share instruments, which leverage pre-existing proximity to InterTubes long-haul fiber nodes and the 1980 county share of U.S. urban college population as shares, and both Chinese and rest-of-the-world data center revenue growth as shifts. The IV estimates show positive effects on total employment, data-processing employment, construction employment, establishments, house prices, and electricity prices at different horizons after data center growth. We also find positive effects on tax returns, adjusted gross income, and wages, while annual payroll responds less robustly. The results suggest that data centers create measurable local activity, increase house prices, and affect local electricity markets through higher prices.

***Keywords*:** Data centers; Artificial intelligence; Digital infrastructure, Local effects

*Authors' correspondence: falvare@uchicago.edu, david.argente@yale.edu, joyce.chow@yale.edu, and diana.vanpatten@yale.edu.

1 Introduction

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

The local stakes are high. A data center can look attractive to local governments because it brings a large capital project, construction activity, a property-tax base, and specialized activity in data-processing and related services. At the same time, data centers are energy-intensive and spatially concentrated. They can require expansions, transmission upgrades, and long-term power contracts. They may also use scarce land and water, while creating fewer permanent jobs than a traditional manufacturing plant with similar capital expenditure. The result is a practical policy question: what do counties actually gain, and what costs do they face, upon data centers entry?

This question has become more urgent with the rise of AI. Data centers accounted for around 1.5 percent of global electricity consumption in 2024, or about 415 TWh, and the United States accounted for the largest share of that global data center electricity use (International Energy Agency, 2025). In the United States, the Department of Energy reports that data center electricity use rose from 58 TWh in 2014 to 176 TWh in 2023. This 300% increase represented about 4.4 percent of U.S. electricity consumption, with large uncertainty but substantial projected growth through 2028 (U.S. Department of Energy, 2024; Shehabi et al., 2024). While these are national totals, impacts are highly local. Data centers are spatially clustered, so their effects on labor markets, utilities, and prices are concentrated on exposed counties.

This paper studies whether counties more exposed to data center growth experience changes in employment, establishments, payroll, income, house prices, and electricity prices. The analysis considers long differences from 1995 to years 2005, 2010, 2015, and 2020. We leverage facility-level data center activity from the 451 Research Datacenter Knowledge Base (DCKB), a proprietary S&P Global/451 Research database with site-level information on facilities, capacity, and annualized revenue. We also measure local economic outcomes from County Business Patterns (CBP), IRS Statistics of Income county files, county-level house prices, and county-level electricity prices constructed from EIA Form 861 data.

The main identification challenge is that data centers do not locate randomly. Firms choose locations with favorable power prices, grid capacity, land, tax policy, climate, disaster risk, and fiber connectivity. We address this concern with a two-instrument shift-share design. The first instrument interacts a county's pre-existing proximity to a long-haul fiber node with growth in Chinese data center revenue. The second interacts a county's 1980 share of national urban college population with rest-of-world data center revenue growth, outside the United States and China.

The first stage is strong across horizons.[1] The IV estimation then shows that data center growth has positive local effects on employment and number of establishments. The most persistent labor-market response occurs in data-processing employment, while construction employment is strongest at short horizons, consistent with requirements to build the centers themselves. We also find positive effects on tax returns, adjusted gross income, and wages, although annual payroll responds less robustly. Finally, electricity-price and house-price effects are positive and highly significant across horizons, indicating that data centers place measurable pressure on local energy and real estate markets. In sum, we document that data centers generate localized employment, increase the number of establishments, increase house prices and lead to aggregate income gains, while also raising local electricity prices.

Related literature. This paper is related to work on digital infrastructure and local labor markets. Broadband and internet infrastructure can affect wages, productivity, and firm location, but the effects depend on technology, adoption, and local absorptive capacity (Forman et al., 2012; Kolko, 2012; Akerman et al., 2015). Data centers, however, differ from household broadband access. They are spatially concentrated capital projects, and their effects operate through construction activity, specialized operations, electricity demand, tax base, and local supply chains.

The paper also relates to research on large plant openings and place-based economic development. Greenstone et al. (2010) use runner-up locations for large plant openings to study agglomeration spillovers. Moretti (2010) studies local multipliers, while Busso et al. (2013) and Glaeser and Gottlieb (2008) discuss the incidence and efficiency of place-based policies. Data centers are relevant in this context because local governments often compete for them, but their labor and power intensity differ from traditional manufacturing plants.

Finally, the paper relates to research on data center electricity use. Shehabi et al. (2016) estimate U.S. data center energy use, Masanet et al. (2020) reassess global data center energy demand in light of efficiency improvements, and Shehabi et al. (2024) report recent U.S. data center energy trends and future scenarios. We study a different margin: how the local arrival and expansion of data centers is associated with county-level employment, establishments, income, house prices, and electricity prices. Contemporaneous work by Yue and Zeng (2026) studies the local effects of data-center entry using a difference-in-differences

[1] In the total-employment baseline specifications, first-stage F-statistics range from 786.1 to 907.0.

design. Entry, however, is likely to be an endogenous outcome: data centers are sited in places with favorable infrastructure, energy access, permitting conditions, tax incentives, expected growth, and other local advantages (e.g. JLL 2026; CBRE 2026, 2025). Our instruments are based on a shift-share design, aiming to relax this concern by using historical local exposure interacted with global data-center revenue shocks. This difference in identification, as well as our focus on cumulative data-center scale rather than first entry, likely explains differences in magnitudes.

2 Data

Our analysis combines facility-level information on data centers with county-level measures of employment, business activity, income, house prices, and electricity prices. The empirical unit is a U.S. county observed over time. We use the facility data to measure where data centers operate and how their scale evolves, and we use public county-level sources to measure the local economic conditions that may respond to this expansion.

Data centers. The data center panel comes from the 451 Research Datacenter Knowledge Base (DCKB). DCKB is a proprietary facility-level database maintained by S&P Global Market Intelligence. It covers more than 12,960 data center facilities across over 2,700 providers and 131 countries, and reports detailed information on facility location, ownership, capacity, and annualized revenue (S&P Global Market Intelligence, 2026).

We map each U.S. facility to a county using its geographic coordinates and county boundary shapefiles. We then aggregate facility-level information to the county-year level. The resulting panel records whether a county has an operating data center in a given year, the number of facilities, and the total annualized revenue associated with those facilities. Our main measure of data center activity is cumulative annualized revenue. Namely, for county c and horizon h, with end year $t_h = 1995 + h$, we define

$$\Delta_h R_c = \log\left(\sum_{t \leq t_h} R_{ct}\right) - \log\left(\sum_{t \leq 1995} R_{ct}\right), \tag{1}$$

where R_{ct} is annualized data center revenue in county c and year t.[2] The cumulative measure captures the scale of data center activity that has arrived in a county by a given horizon. We also report specifications using cumulative facility counts, an indicator for whether the county has a data center by the end year, and an ever-treated indicator.

[2]For variables that can be zero, the empirical implementation uses the inverse hyperbolic sine transformation.

The same DCKB data are also used to construct the aggregate shifts in the shift-share design. We build annual measures of data center revenue for the United States, China, and the rest of the world outside the United States and China. The baseline instruments interact historical county exposure with growth in Chinese data center revenue and with revenue growth in the rest of the world outside the United States and China.

County Business Patterns. We measure local business activity using County Business Patterns (CBP). The main CBP outcomes are total employment, total establishments, and annual payroll. These variables provide a broad picture of county-level economic activity in the private sector. We also construct two sectoral employment measures that are especially relevant for data centers. The first is data-processing employment, which captures activity in industries directly related to hosting, processing, and related digital infrastructure services. In the SIC period, we use SIC 7374, Computer Processing and Data Preparation and Processing Services. In the NAICS period, we use NAICS 518210, Data Processing, Hosting, and Related Services. The second is construction employment, defined using SIC 15 and NAICS 236. This outcome is intended to capture the local build-out margin associated with constructing large facilities.[3]

Income and tax outcomes. We use county-level IRS Statistics of Income files to measure broader local income responses. These data provide the number of tax returns, adjusted gross income, and wages at the county-year level. These outcomes are useful because employment gains around data centers need not imply broad income gains for local residents. Some jobs may be temporary, specialized, or filled by workers outside the county. The IRS outcomes therefore allow us to ask whether data center expansion is visible in broader county-level income aggregates, beyond employment and establishments.

House prices. We measure local housing-market responses using the county-level House Price Index (HPI) from the Federal Housing Finance Agency (FHFA). Specifically, we use the FHFA annual county HPI workbook, which provides repeat-sales house price indexes at the county-year level. The HPI outcome is the percent change in the index relative to the 1995 baseline.

[3]Because the industry classification system changes from SIC to NAICS during the sample, the sector-specific series should be interpreted as approximate, not as perfectly harmonized industry measures. This caveat is less important for the aggregate CBP outcomes, which are measured consistently at the county level.

Electricity prices. We measure electricity prices using a county-level panel constructed from EIA Form 861. The underlying EIA files report utility-level revenue, electricity sales, and customer counts by sector, as well as service-territory information linking utilities to counties. The county-level price is constructed by aggregating revenue and MWh across utilities serving a county and then dividing revenue by sales. Our main electricity outcome is the total price. Conceptually, this variable captures the average price paid by electricity consumers in a county.

Historical shares The shift-share design uses two predetermined county-level exposure measures. The first is proximity to long-haul fiber infrastructure. Data centers require reliable high-capacity connectivity, and historical fiber routes created persistent geographic advantages for some locations. We measure this exposure as negative log distance to the nearest InterTubes long-haul fiber node. The node locations come from the Internet Atlas and InterTubes projects, which document long-haul fiber links, nodes, conduits, and conduit sharing among U.S.-based providers (Durairajan et al., 2013, 2015). We assign FIPS codes to node locations and use the NBER county distance database to compute county-to-node distances (Roth, 2024). The measure is predetermined relative to the post-1995 expansion studied in the paper. In constructing it, we exclude six nodes for which we were unable to verify clear pre-2000 tube or conduit evidence. The second exposure measure is the county's 1980 share of the national urban college population. This variable is constructed from historical Census county data using total population, urban population, and college population in 1980 (U.S. Census Bureau, 1998; Schroeder et al., 2025). The motivation is that counties with larger concentrations of urban college-educated population in 1980 were more exposed to the early human-capital and urban-demand conditions that shaped the geography of digital infrastructure. Because the share is measured fifteen years before the 1995 baseline, it is predetermined with respect to the subsequent growth in data center revenue that we use as the shift.

3 Industry background

A fast-growing, concentrated industry Data centers house servers, storage, and networking equipment, together with the power and cooling systems needed to keep that equipment running. Their growth reflects several demand forces: the migration of enterprise computing to the cloud, software-as-a-service, streaming and digital public services, and now AI training and inference. To better understand this industry's dynamics, we leverage data center facility-level data. Namely, Figure 1 considers the total number of data centers globally and across time. The figure shows a sharp acceleration after the mid-2000s, with especially rapid growth in Asia-Pacific, the United States and Canada, and Europe. The U.S. market is also highly concentrated. Figure 2 maps state-level data center counts and shows the outsized role of Virginia and Texas.

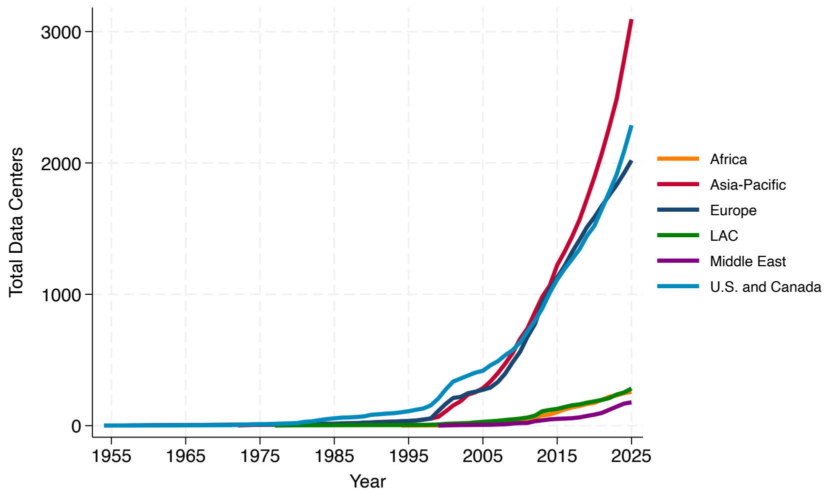


Figure 1: Growth in data centers by subregion

Notes: The figure plots cumulative data center counts by subregion across time.

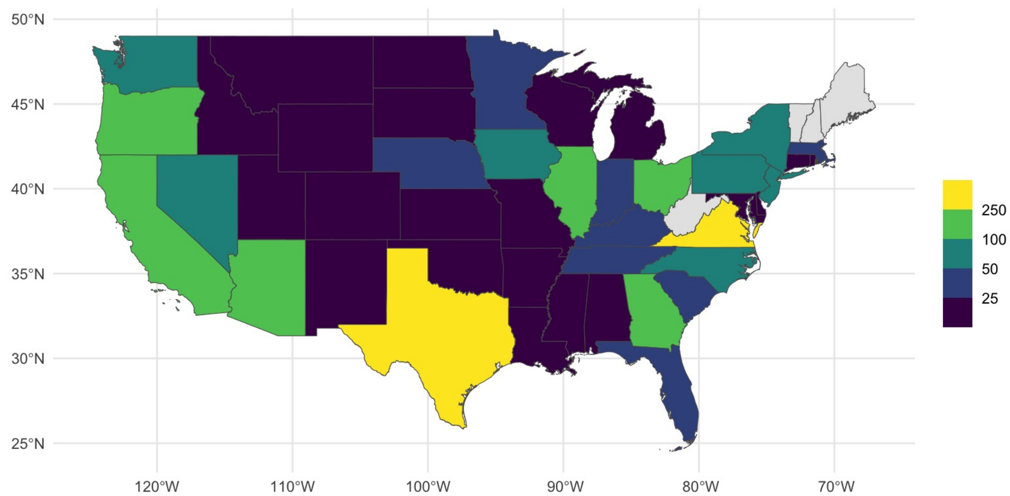
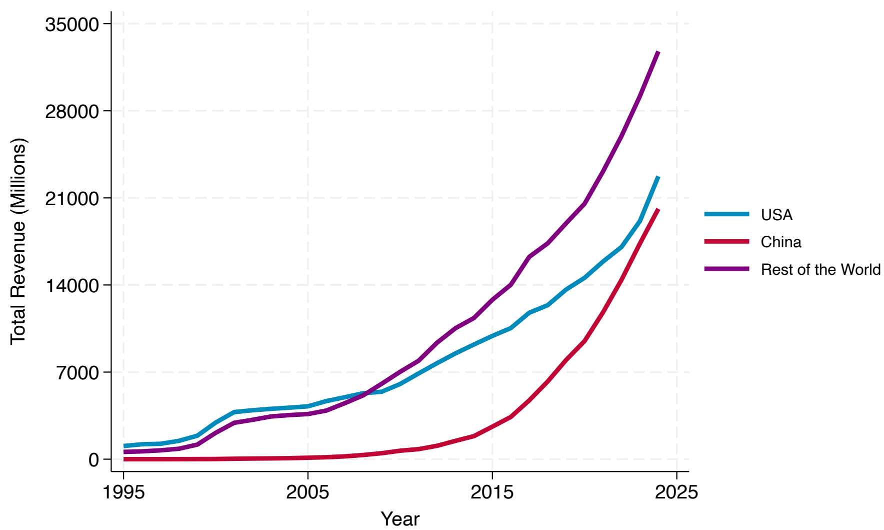


Figure 2: Number of data centers by state

Notes: The figure reports state-level facility counts in the contiguous United States.

Revenue, capacity and construction costs A useful way to think about a data center is as a factory for computation. Its output is compute, storage, and network services. Its inputs are land, structures, power equipment, cooling, servers, specialized labor, and connectivity. Industry capacity is therefore measured in several ways. Figure 3 shows the growth in revenue from 1995 onwards. Revenue in China is zero prior to 1995, and very low elsewhere.

Figure 3: Cumulative total revenue

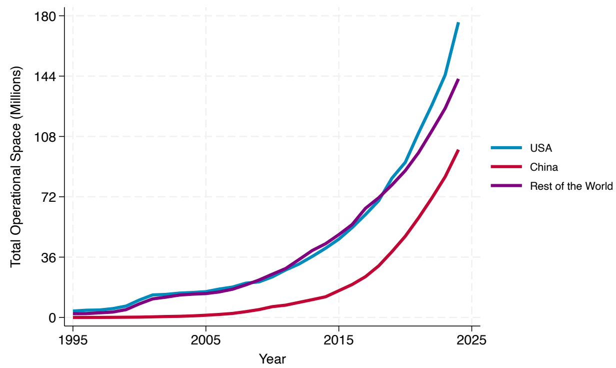


Notes: The figure shows the total cumulative revenue (in millions of US dollars) over time.

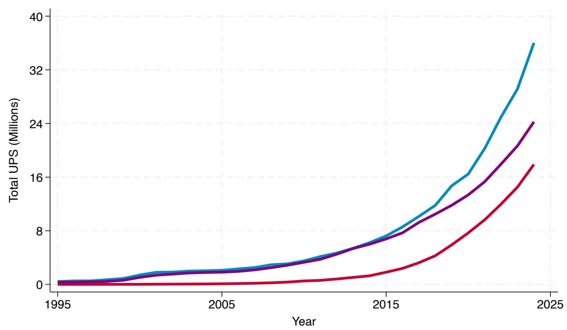
Figure 4 shows that capacity measures grew sharply after 1995, especially after 2015. Namely, Panels (a) descrives the total operational space required by the data centers. Panel considers the total count of racks; a measure of the number of equipment cabinets, and Panel (c) relies on an Uninterruptible Power Supply (UPS) measure, which depedns on the electricity capacity available to IT equipment and is one of the most common industry measures of data center scale.

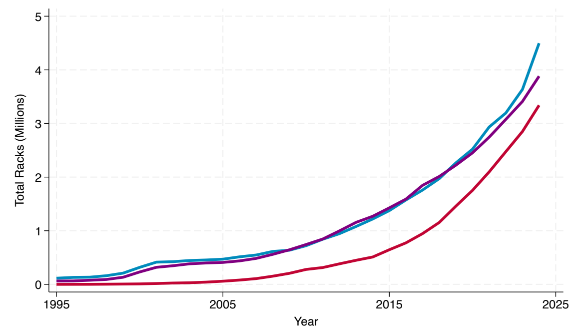
Critical capacity, the electrical and mechanical infrastructure that supports IT load, accounts for 53% of construction costs (Cushman & Wakefield, 2025). This helps explain why data centers can generate large construction demand and large utility-system requirements even when permanent employment is modest. Land is another source of geographic differentiation. There are large differences in land prices across selected data center markets. Established markets such as Chicago, Phoenix, and Virginia command large land premia relative to several emerging markets, like Iowa, Indiana, and Minneapolis (Cushman & Wakefield, 2025). Those premia reflect the same factors that make data center location endogenous: fiber, power availability, permitting, proximity to demand, and the ability to

Figure 4: Cumulative capacity growth of data centers



(a) Cumulative operational space





(b) Cumulative UPS

(c) Cumulative total racks

Notes: The figure shows the cumulative dynamics of data center outcomes. Panel (a) considers total operational space (in millions of square feet), Panel (b) considers the cumulative UPS consumed, and Panel (c) displays the cumulative number of racks.

expand.

Electricity, cooling, and efficiency Electricity is central to data center operations. In a typical facility, power reaches IT equipment through switch gear, uninterruptible power supplies, and power distribution units. Cooling systems and other support loads sit alongside the servers, storage devices, and network equipment. This architecture explains why the same facility can affect both local employment and local electricity prices.

Location choice Data center location depends on power prices, grid capacity, fiber access, latency, land, climate, water, permitting, taxes, and disaster risk. These forces make naive comparisons of data center and non-data center counties hard to interpret. They also motivate the shift-share design proposed below to account for endogenous site choice.

4 Empirical strategy

The baseline second-stage equation for horizon h is

$$\Delta_h Y_c = \beta_h \widehat{\Delta_h R_c} + \varepsilon_{ch}, \tag{2}$$

where $\Delta_h Y_c$ is the change in the log outcome from 1995 to year $1995 + h$, and $\Delta_h R_c$ is the change in log cumulative data center annualized revenue. For the HPI, $\Delta_h Y_c$ is the percentage change in the index relative to 1995. The first stage is

$$\Delta_h R_c = \pi_{1h} Z^F_{ch} + \pi_{2h} Z^U_{ch} + v_{ch}. \tag{3}$$

The two excluded instruments are

$$Z^F_{ch} = [-\log(\text{distance to nearest fiber node}_c)] \times \Delta_h R^{China}, \tag{4}$$

$$Z^U_{ch} = (1980 \text{ urban college share}_c) \times \Delta_h R^{ROW}, \tag{5}$$

where $\Delta_h R^{China}$ is the change in log data center annualized revenue in China. Finally, $\Delta_h R^{ROW}$ is the corresponding rest-of-world change, excluding the United States and China. In the data center time series, both China and the rest-of-the-world have virtually no data center revenue before the 1995 baseline, which makes these shocks post-baseline expansions.

The empirical design follows the logic of shift-share or Bartik instruments (Bartik, 1991), where identification can be framed as relying on the exogeneity of predetermined shares (Goldsmith-Pinkham et al., 2020) *or* on the quasi-randomness of shocks (Borusyak et al., 2022). While our shares are predetermined, our shocks are arguably quasi-random from individual U.S. counties' perspective, and in the spirit of the *China shock* literature (Autor et al., 2013).

5 Descriptive statistics for main outcomes

Table 1 reports summary statistics for the 1995 to 2020 long differences. Counties are highly heterogeneous. The mean change in data-processing employment is large relative to the median county, but many counties have no data center exposure. The data center treatment is sparse: by 2020, 7.6 percent of counties have positive data center presence over the horizon, while the mean change in log cumulative data center revenue is 0.190.

Table 2 shows how treatment exposure grows across horizons. The average data center revenue and facility-count exposure rises from 2005 to 2020. The share of counties with

Table 1: Summary statistics for 1995 to 2020 long differences

Variable	Obs.	Mean	Std. dev.
% Δ employment	3,173	0.212	0.858
% Δ data-processing employment	3,173	0.525	1.627
% Δ construction employment	3,173	-1.430	1.530
% Δ annual payroll	3,173	0.969	1.039
% Δ establishments	3,173	0.091	0.488
% Δ tax returns	3,130	0.338	0.297
% Δ adjusted gross income	3,130	0.958	0.358
% Δ wages	3,130	0.861	0.340
% Δ house price index	2,046	1.093	0.441
% Δ electricity price	2,954	0.038	0.022
% Δ cumulative data center revenue	3,188	0.190	0.869
% Δ cumulative data center facilities	3,188	0.130	0.524
Data center present over horizon	3,188	0.076	0.264

Notes: Variables are county-level long differences from 1995 to 2020. Rows labeled as percent changes are inverse-hyperbolic-sine differences in the implementation, which approximate log differences for positive values and retain observations with zeros. The house price index row is the percentage change in the HPI relative to 1995.

positive data center presence during the horizon rises from 3.0 percent in 2005 to 7.6 percent in 2020.

Table 2: Mean treatment exposure by horizon

	2005	2010	2015	2020
Cumulative data center revenue	0.080	0.115	0.172	0.190
	(0.554)	(0.658)	(0.815)	(0.869)
Cumulative data center facilities	0.043	0.066	0.106	0.130
	(0.274)	(0.346)	(0.459)	(0.524)
Data center present over horizon	0.030	0.044	0.065	0.076
	(0.170)	(0.206)	(0.246)	(0.264)

Notes: The table reprts mean treatment by horizon. Standard deviations are in parenthesis. The cumulative outcomes report inverse-hyperbolic-sine long differences in cumulative data center revenue and facility counts. The deta center presence outcome reports an indicator equal to one when a county has positive data center presence during the horizon.

6 First stage

Table 3 reports the first stage for the main endogenous treatment. Both instruments are highly predictive. The first-stage F-statistics range from 786.1 to 907.0 across horizons in the total-employment sample. The coefficient on the 1980 urban college share interacted with rest-of-world growth is negative in the long difference with respect to 2005 because the rest-of-world revenue shock is negative over 1995 to 2005 in the constructed data center aggregate. In later horizons, however, it is positive.

Table 3: First stage for cumulative data center annualized revenue

End year	$-$log distance × China shock		1980 urban college share × ROW shock		First-stage F	R^2	Obs.
	Coef.	SE	Coef.	SE			
2005	0.010***	(0.003)	-569.681***	(14.811)	786.11	0.335	3,130
2010	0.012***	(0.002)	118.561***	(3.071)	811.24	0.342	3,131
2015	0.016***	(0.002)	126.021***	(3.131)	904.22	0.366	3,131
2020	0.016***	(0.002)	130.279***	(3.238)	906.95	0.367	3,129

Notes: The dependent variable is the horizon-specific change in log cumulative data center annualized revenue. Robust standard errors are in parentheses. The China and ROW shocks are long differences in data center annualized revenue outside the United States, with ROW excluding both the United States and China. Significance: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

7 Main IV results

Employment and establishments Table 4 reports the IV estimates related to employment and number of establishments. Total employment increases at most horizons, with statistically significant effects in 2005, 2010, and 2020. The 2020 coefficient is 0.039, implying that a one-unit increase in cumulative data center revenue is associated with a 3.9 percent increase in total employment between 1995 and 2020. This effect is not mechanically driven by data center jobs themselves. When we instead use total county employment excluding data-processing employment, the coefficient is 0.038. This similarity is expected, since data-processing employment accounts for less than one percent of total county employment.

The more direct labor-market response appears in data-processing employment. The estimates imply that a one-unit increase in cumulative data center revenue raises data-processing employment by roughly 27 percent in 2015 and 29 percent in 2020. Construction employment follows a different time profile. The largest coefficients appear in 2005 and 2010, consistent with a build-out phase in which the local labor demand created by data centers is concentrated around construction and installation. Establishments also rise at every horizon, with coefficients between 0.026 and 0.062.

House and electricity prices House prices increase as well. As shown in Table 4, HPI coefficients are positive and statistically significant in every horizon, with a 2020 coefficient of 0.177. Electricity prices also increase in all horizons. The 2020 coefficient is 0.009, and the electricity-price estimates are statistically significant at the 1 percent level in all baseline horizons. The table therefore points to both benefits and costs. Data centers generate measurable local activity, especially in sectors directly connected to their construction and operation, and they also place upward pressure on house prices and electricity prices.

Table 4: Main IV estimates: employment, establishments, house prices, and electricity prices

	End year			
Outcome	2005	2010	2015	2020
Employment	0.022*	0.063***	0.014	0.039***
	(0.012)	(0.020)	(0.010)	(0.011)
Data-processing employment	0.092	-0.119	0.265***	0.294***
	(0.067)	(0.073)	(0.048)	(0.053)
Construction employment	0.600***	0.473***	0.052***	0.071***
	(0.117)	(0.075)	(0.019)	(0.019)
Establishments	0.026***	0.034***	0.046***	0.062***
	(0.007)	(0.008)	(0.007)	(0.008)
House price index	0.239***	0.081***	0.124***	0.177***
	(0.039)	(0.019)	(0.021)	(0.028)
Electricity price	0.007***	0.008***	0.008***	0.009***
	(0.002)	(0.002)	(0.001)	(0.001)
Observations, employment sample	3,130	3,131	3,131	3,129
First-stage F, employment sample	786.11	811.24	904.22	906.95

Notes: Each cell reports a separate 2SLS coefficient of the listed long-difference outcome on the horizon-specific change in log cumulative data center annualized revenue. Robust standard errors are in parentheses. All specifications use the two excluded instruments reported in Table 3. Rows are log-change approximations, with inverse-hyperbolic-sine differences used for variables that can be zero. The HPI outcome is the percentage change in the house price index relative to 1995. Electricity-price regressions use 2,986, 2,997, 2,975, and 2,954 observations across the four horizons; HPI regressions use 2,036 observations in each horizon. Significance: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Payroll, income, and wages Table 5 reports outcomes for annual payroll, tax returns, adjusted gross income, and wages. The estimates show positive effects on several broad income aggregates. Tax returns, adjusted gross income, and wages are positive and statistically significant in every horizon. The 2020 coefficients imply increases of about 5.9 percent in tax returns, 8.2 percent in adjusted gross income, and 5.9 percent in wages. Annual payroll is positive and significant in 2005 and 2010, but not in the longer horizons.

Table 5: Main IV estimates: payroll, income, and wages

	End year			
Outcome	2005	2010	2015	2020
Annual payroll	0.046***	0.049**	-0.005	0.017
	(0.018)	(0.021)	(0.013)	(0.013)
Tax returns	0.021***	0.050***	0.058***	0.059***
	(0.007)	(0.010)	(0.009)	(0.010)
Adjusted gross income	0.021***	0.055***	0.065***	0.082***
	(0.008)	(0.013)	(0.009)	(0.011)
Wages	0.020***	0.040***	0.054***	0.059***
	(0.008)	(0.010)	(0.009)	(0.009)
Observations, payroll sample	3,130	3,131	3,131	3,129
First-stage F, payroll sample	786.11	811.24	904.22	906.95

Notes: Each cell reports a separate 2SLS coefficient of the listed long-difference outcome on the horizon-specific change in log cumulative data center annualized revenue. Robust standard errors are in parentheses. All specifications use the two excluded instruments reported in Table 3. Rows are log-change approximations, with inverse-hyperbolic-sine differences used for variables that can be zero. Observation counts differ slightly across IRS and average-pay outcomes because of data availability. Significance: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

8 Alternative treatment definitions

Table 6 reports 1995 to 2020 estimates using alternative measures of data center exposure. The first column is the baseline continuous revenue treatment. The second uses cumulative data center facilities. The third uses an indicator for whether the county has a data center over the 25-year horizon. The fourth uses an ever-data center indicator. The latter is included as a diagnostic because it is defined over the full available panel rather than only by the horizon.

The qualitative pattern is stable across treatment definitions. Data-processing employment, construction employment, establishments, house prices, and electricity prices are positive in all columns. The binary specifications have larger coefficients because the treatment changes from a continuous log revenue measure to an indicator. These estimates should therefore be interpreted as effects of moving from no data center presence to data center presence, not as elasticities.

Table 6: Alternative endogenous treatment measures, 1995 to 2020

Outcome	Log DC revenue	Log DC facilities	Ever DC
Employment	0.039***	0.065***	0.175***
	(0.011)	(0.019)	(0.049)
Data-processing employment	0.294***	0.490***	1.327***
	(0.053)	(0.091)	(0.178)
Construction employment	0.071***	0.130***	0.278***
	(0.019)	(0.032)	(0.085)
Establishments	0.062***	0.102***	0.288***
	(0.008)	(0.015)	(0.029)
Electricity price	0.009***	0.015***	0.039***
	(0.001)	(0.002)	(0.005)
House price index	0.177***	0.314***	0.772***
	(0.028)	(0.053)	(0.124)
Annual payroll	0.017	0.031	0.066
	(0.013)	(0.021)	(0.062)
Tax returns	0.059***	0.094***	0.278***
	(0.010)	(0.016)	(0.032)
Adjusted gross income	0.082***	0.136***	0.369***
	(0.011)	(0.019)	(0.037)
Wages	0.059***	0.098***	0.263***
	(0.009)	(0.017)	(0.034)
Observations, CBP outcomes	3,129	3,129	3,129
Observations, IRS outcomes	3,130	3,130	3,130
Observations, electricity	2,954	2,954	2,954
Observations, HPI	2,036	2,036	2,036
First-stage F, CBP sample	906.95	776.48	437.91
First-stage F, IRS sample	907.29	776.77	438.09

Notes: Each cell reports a separate 2SLS coefficient for the 1995 to 2020 outcome. The first two columns use horizon-specific treatments. The ever-data-center column uses treatment status over the full available panel and is included as a diagnostic specification. Rows are log-change approximations, with inverse-hyperbolic-sine differences used for variables that can be zero. The house price index row is the proportional change in the FHFA county-level HPI relative to 1995. Robust standard errors are in parentheses. Significance: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

9 Robustness

Table 7 reports 1995 to 2020 robustness checks for the main continuous revenue treatment. The first column repeats the baseline. The second absorbs state fixed effects and clusters by state. With state fixed effects, the positive effects on data-processing employment, establishments, and house prices remain statistically significant. The total-employment coefficient falls from 0.039 to 0.026 and is no longer statistically significant. Construction and electricity-price estimates are also no longer statistically significant with state fixed effects.

Table 7: Robustness for 1995 to 2020 estimates

Outcome	Baseline	State fixed effects
Employment	0.039***	0.026
	(0.011)	(0.016)
Data-processing employment	0.294***	0.246***
	(0.053)	(0.073)
Construction employment	0.071***	0.003
	(0.019)	(0.033)
Establishments	0.062***	0.063***
	(0.008)	(0.008)
Electricity price	0.009***	0.001
	(0.001)	(0.002)
House price index	0.177***	0.136***
	(0.028)	(0.034)
Annual payroll	0.017	0.020
	(0.013)	(0.020)
Tax returns	0.059***	0.042***
	(0.010)	(0.012)
Adjusted gross income	0.082***	0.067***
	(0.011)	(0.011)
Wages	0.059***	0.056***
	(0.009)	(0.011)
Observations, CBP outcomes	3,129	3,129
Observations, IRS outcomes	3,130	3,130
Observations, electricity	2,954	2,954
Observations, HPI	2,036	2,035
First-stage statistic, CBP sample	906.95	46.25
First-stage statistic, IRS sample	907.29	46.27

Notes: Each cell reports a separate 2SLS coefficient for the 1995 to 2020 outcome on the change in log cumulative data-center annualized revenue. The state fixed effects column absorbs state indicators and clusters standard errors by state; the first-stage statistic in that column is the Kleibergen-Paap Wald statistic reported in the log. Rows are log-change approximations, with inverse-hyperbolic-sine differences used for variables that can be zero. The house price index row is the proportional change in the FHFA county-level HPI relative to 1995. Robust standard errors are in parentheses in the baseline column; state-clustered standard errors are in parentheses in the state fixed effects column. Significance: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

10 Conclusion

This paper studies the local effects of data center expansion using county-level long differences and two historical shift-share instruments. The instruments combine pre-1995 digital infrastructure and 1980 human-capital geography with global data center growth shocks. The estimates show positive effects on total employment, data-processing employment, establishments, construction employment during early build-out horizons, house prices, and electricity prices. We also find positive effects on tax returns, adjusted gross income, and wages. Annual payroll responds less robustly.

The policy implication is therefore mixed. Data centers create economic activity, especially in directly related sectors and during construction, and they are associated with larger county-level income aggregates. They also raise electricity prices and are associated with higher house prices, which may benefit property owners while increasing costs for renters and prospective homebuyers. Counties considering data center projects should therefore weigh employment, establishment, and aggregate income gains against pressures on electricity markets and local housing affordability.

References

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

Autor, D.H., Dorn, D., Hanson, G.H., 2013. The china syndrome: Local labor market effects of import competition in the united states. American Economic Review 103, 2121–68. URL: https://www.aeaweb.org/articles?id=10.1257/aer.103.6.2121, doi:10.1257/aer.103.6.2121.

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

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

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

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

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

Cushman & Wakefield, 2025. U.S. data center development cost guide. Industry report. Data-center development cost and land cost benchmarks.

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.

Durairajan, R., Ghosh, S., Tang, X., Barford, P., Eriksson, B., 2013. Internet atlas: A geographic database of the internet, in: Proceedings of ACM HotPlanet.

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

Glaeser, E.L., Gottlieb, J.D., 2008. The economics of place-making policies. Brookings Papers on Economic Activity , 155–253.

Goldsmith-Pinkham, P., Sorkin, I., Swift, H., 2020. Bartik instruments: What, when, why, and how. American Economic Review 110, 2586–2624.

Greenstone, M., Hornbeck, R., Moretti, E., 2010. Identifying agglomeration spillovers: Evidence from winners and losers of large plant openings. Journal of Political Economy 118, 536–598.

International Energy Agency, 2025. Energy and AI. IEA report. Analysis of electricity demand from data centres and artificial intelligence.

JLL, 2026. 2026 global data center outlook. URL: https://www.jll.com/en-us/insights/market-outlook/data-center-outlook. jones Lang LaSalle Research.

Kolko, J., 2012. Broadband and local growth. Journal of Urban Economics 71, 100–113.

Masanet, E., Shehabi, A., Lei, N., Smith, S., Koomey, J., 2020. Recalibrating global data center energy-use estimates. Science 367, 984–986.

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

Roth, J., 2024. County distance database. National Bureau of Economic Research data file.

Schroeder, J.P., Van Riper, D., Manson, S., Knowles, K., Kugler, T., Roberts, F., Ruggles, S., 2025. IPUMS national historical geographic information system: Version 20.0 [dataset].

Shehabi, A., Smith, S.J., Hubbard, A., Newkirk, A., Lei, N., Siddik, M.A.B., Holecek, B., Koomey, J.G., Masanet, E.R., Sartor, D.A., 2024. 2024 United States Data Center Energy Usage Report. Technical Report LBNL-2001637. Lawrence Berkeley National Laboratory.

Shehabi, A., Smith, S.J., Sartor, D.A., Brown, R.E., Herrlin, M., Koomey, J.G., Masanet, E.R., Horner, N., Azevedo, I.L., Lintner, W., 2016. United States Data Center Energy Usage Report. Technical Report LBNL-1005775. Lawrence Berkeley National Laboratory.

S&P Global Market Intelligence, 2026. 451 research datacenter knowledgebase. Product documentation and database accessed through Yale subscription.

U.S. Census Bureau, 1998. Usa counties: Historical county data files. Historical county data used for 1980 population, urban population, and education tabulations.

U.S. Department of Energy, 2024. DOE releases new report evaluating increase in electricity demand from data centers. Press release summarizing the 2024 United States Data Center Energy Usage Report.

Yue, D., Zeng, Y., 2026. The local economic effects of data center entry. Working paper, SSRN.