The Local Economic Effects of Data Center Entry

Daniel Yue* Yiyang Zeng*

March 2026

Abstract

The explosive growth of artificial intelligence has triggered an unprecedented wave of data center investment, yet systematic evidence on how these facilities affect local economies remains scarce. We study the local economic effects of data center entry in the United States using comprehensive facility-level data merged with county-level economic outcomes. Employing a Callaway-Sant'Anna difference-in-differences framework that accounts for staggered treatment timing, we find that data center activation increases local employment by 3.5%, total wages by 5.0%, business establishments by 4.7%, and median household income by 1.9%, while reducing unemployment rates and stimulating building permit activity. These effects are concentrated in metropolitan areas, where agglomeration economies amplify multiplier effects, while non-metro counties experience negligible gains. Big Tech facilities generate larger effects than third-party data centers, and counties with early first-entry timing experienced stronger effects than those with recent first-entry timing. We also find that counties receiving additional data centers within five years of first entry exhibit larger effects than counties with isolated entries, suggesting that data center clustering amplifies local economic benefits. Finally, we examine electricity price effects, finding that identification is complicated by utility service territories spanning multiple counties. In areas with cleaner identification (fewer shared utilities), we detect positive price effects, suggesting data center demand raises costs for local ratepayers. Our findings inform ongoing policy debates about data center expansion and its local economic consequences.

Keywords: data centers, artificial intelligence, innovation and technology policy

*Scheller College of Business, Georgia Institute of Technology
daniel.yue@scheller.gatech.edu yiyang.zeng@scheller.gatech.edu

1 Introduction

The rapid rise of artificial intelligence has fueled explosive growth in data centers, which now form the backbone of training large-scale models. Figure 1 illustrates how AI-related spending compares to prior major capital expenditure waves. These estimates place spending at roughly 1.2% of U.S. GDP—already larger than the fiber-optic boom of the 2000s and rivaling the railroad build-out of the nineteenth century. This scale puts data centers among the largest waves of capital investment in American history.

Yet despite the size of this surge, we know surprisingly little about its local consequences. Media accounts highlight ballooning electricity demand, rising water usage, and costly subsidies for firms such as Amazon, Microsoft, and Google. Research in environmental sciences underscores the sustainability tensions of this expansion: U.S. data centers have met rising demand without proportional increases in electricity use as a result of efficiency gains (Masanet et al., 2020), yet they continue to impose heavy environmental costs in both power and water consumption (Shehabi et al., 2018; Siddik et al., 2021).

In economics and management, however, systematic empirical evidence on how data center entry affects local jobs, incomes, and public finances remains almost entirely absent. Some studies examine energy efficiency in production (Park et al., 2023) and the entry strategies of third-party data centers (Fang and Greenstein, 2025), but little is known about the broader economic effects of data centers. Anecdotal accounts from Virginia—the nation's largest data center hub—suggest that while data centers directly employ only a small share of the workforce, they create highly productive jobs and exert strong regional spillovers (Mullin, 2023). Given the growing importance and the limited evidence on this topic, it is of interest to researchers and policymakers to study the economic role of AI-driven data centers, including their effects on labor markets, regional development, and ultimately the geography of economic growth.

The stakes are high. If AI-related investment proves durable, it could reshape local economies in ways reminiscent of past industrial transformations. If it proves fleeting, many communities risk being left with stranded assets and depleted budgets. For policymakers, this uncertainty raises difficult questions about incentive designs. If the investments ultimately generate no return, subsidies represent a deadweight fiscal loss. If the returns exist but arrive only with a long lag, governments may need to bridge short-term costs until the benefits materialize. Meanwhile, data center expansion has provoked local resistance. Communities in Virginia, Iowa, California, and elsewhere have mobilized against new projects, citing concerns about noise, land use, and resource strain.

This paper provides the first comprehensive empirical analysis of the local economic

effects of data center entry. Using detailed facility-level data from Aterio merged with county-level outcomes from the Quarterly Census of Employment and Wages (QCEW) and other sources, we examine how data center activation affects employment, wages, establishments, median household income, unemployment rates, and building permits. We employ the Callaway and Sant'Anna (2021) difference-in-differences estimator, which properly accounts for staggered treatment timing and avoids the negative weighting problems that plague traditional two-way fixed effects designs.

Our main findings can be summarized as follows. First, data center entry generates positive and statistically significant effects on local economic activity: employment increases by approximately 3.5%, total wages by 5.0%, business establishments by 4.7%, and median household income by 1.9%. We also observe declines in unemployment rates and increases in building permits, consistent with construction activity and labor market tightening. These magnitudes are economically meaningful, though smaller than effects documented for large manufacturing plants in the "million-dollar plants" literature. Second, the effects are highly heterogeneous across locations. Metropolitan counties experience employment gains of 4.1%, while non-metro counties show no significant effects—consistent with the importance of agglomeration economies for realizing multiplier effects. Third, Big Tech facilities generate larger effects than third-party data centers, and counties where data centers entered during the cloud computing era experienced stronger impacts than those with more recent entry. Fourth, counties that received additional data centers within five years of their first entry exhibit larger effects than counties with isolated entries, suggesting that clustering amplifies local economic benefits. Fifth, we find evidence that data center entry raises local electricity prices, though identification is complicated by the geographic structure of utility service territories.

The remainder of the paper proceeds as follows. Section 2 reviews related literature and develops our hypotheses. Section 3 describes our data and empirical approach. Section 4 presents results. Section 5 discusses implications and limitations. Section 6 concludes.

2 Literature and Hypotheses

2.1 Related Literature

Our study contributes to two strands of research: the economics of large-scale facility openings and the emerging literature on digital infrastructure.

The "million-dollar plants" literature provides our primary theoretical foundation. Greenstone and Moretti (2004) pioneered the use of runner-up counties as counterfactuals

to study the effects of large manufacturing plants, finding that winning counties experienced significant increases in property values. Greenstone et al. (2010) extended this approach to show that incumbent manufacturing plants in winning counties experienced productivity gains of 12 percent, suggesting positive agglomeration spillovers from new entrants. Moretti (2010) documented substantial local multipliers: each additional job in the tradable sector creates approximately 1.6 additional jobs in local non-tradable services. Moretti (2012) synthesized this evidence to argue that large investments can fundamentally reshape regional economic geography, creating persistent divergence between winning and losing locations. Related work on place-based policy extends these insights to large-scale public investments, such as Kline and Moretti (2014), who study the long-run effects of the Tennessee Valley Authority and find persistent increases in manufacturing employment and agricultural productivity.

These studies share a common insight: large, lumpy investments generate effects that extend well beyond direct employment. The mechanisms include backward linkages to suppliers, forward linkages to customers, labor market pooling, and knowledge spillovers. However, the magnitude of these effects depends critically on the characteristics of the investment, the local economic environment, and the nature of the jobs created.

A parallel literature examines subsidy competition for business investment. Slattery and Zidar (2020) document that state and local governments spend tens of billions of dollars annually on business incentives, with limited evidence on their effectiveness. Slattery (2025) shows that subsidy competition can lead to inefficient spatial allocation of investment. Busso et al. (2013) evaluate federal Empowerment Zone policies targeting economically distressed urban areas, finding modest employment gains from tax credits and block grants. Data centers have become prominent targets for these incentives, with states offering property tax abatements, sales tax exemptions, and direct grants worth hundreds of millions of dollars for large facilities. Whether these subsidies generate commensurate local benefits remains an open question.

Research specifically on data centers is more limited. Fang and Greenstein (2025) examine the location decisions of third-party data centers, finding that proximity to network infrastructure and low electricity prices are key determinants. Pham (2017) provides descriptive evidence on data center employment patterns. However, systematic causal evidence on the local economic effects of data center entry remains scarce. Our study fills this gap by providing the first comprehensive analysis using facility-level data and the Callaway-Sant'Anna difference-in-differences estimator (Callaway and Sant'Anna, 2021), which addresses the negative weighting bias inherent in traditional two-way fixed effects (TWFE) specifications when treatment effects are heterogeneous across cohorts or time (Goodman-Bacon, 2021; Sun and

Abraham, 2021).

2.2 Data Center Industry Background

Data centers are physical facilities that house computer systems, storage equipment, and networking infrastructure required to process, store, and distribute digital information. While the term "cloud computing" suggests an ethereal, location-independent resource, the cloud is fundamentally grounded in massive physical infrastructure: climate-controlled buildings filled with thousands of servers, connected by fiber-optic networks and powered by dedicated electrical substations. Figure 2 shows the geographic distribution of active data centers across U.S. counties, with most facilities concentrated in metropolitan areas.

Physical infrastructure. A modern data center consists of several key components. The server halls contain rows of computing equipment mounted in standardized racks, with each rack drawing 10–50 kilowatts of electricity. Cooling systems—typically a combination of computer room air handlers, chilled water loops, and increasingly liquid cooling for high-density AI workloads—remove the heat generated by computing equipment. Electrical infrastructure includes uninterruptible power supplies (UPS), backup generators, and often dedicated substations connecting directly to the high-voltage grid. Network infrastructure provides connectivity through multiple fiber-optic connections to internet exchange points.

Construction process. Building a data center involves multiple phases spanning 18–36 months for hyperscale facilities. Site selection prioritizes access to reliable power (preferably from multiple independent substations), fiber connectivity, water for cooling, and land for expansion. Construction begins with site preparation and building shell, followed by electrical infrastructure, cooling systems, and finally server deployment. Unlike manufacturing plants that may use prefabricated components, data centers require extensive on-site electrical and mechanical work, creating substantial demand for local construction trades.

Facility types. The industry distinguishes several facility types. *Hyperscale* data centers, operated by major cloud providers (Amazon Web Services, Google Cloud, Microsoft Azure) or large internet companies (Meta, Apple), are massive facilities of 50–500+ megawatts designed for cloud computing and AI training. *Colocation* facilities, operated by third-party providers (Equinix, Digital Realty, CyrusOne), offer space and power to multiple tenants who bring their own equipment. *Enterprise* data centers serve a single organization's internal IT needs. Our analysis captures all types, though the economic effects likely vary by facility characteristics.

Economic footprint. A typical hyperscale data center costs $500 million to $1 billion to construct and employs 30–100 permanent workers in operations. However, the

economic footprint extends beyond direct employment through construction activity, ongoing maintenance and security services, equipment procurement, and the induced spending of high-wage data center workers in the local economy.

2.3 Conceptual Framework

Data centers affect local economies through multiple channels that operate at different time horizons and scales.

Construction phase effects. The most immediate impact comes during construction. As described above, hyperscale facilities require extensive on-site work over 18–36 months, creating demand for local construction workers, electricians, HVAC technicians, and related trades. The temporary nature of construction employment means these effects should appear in the quarters surrounding facility activation and then dissipate.

Operations phase effects. Once operational, data centers employ relatively few workers—typically 30–100 per facility, depending on size and design. However, these positions tend to be high-skill and high-wage: data center technicians, network engineers, security specialists, and facilities managers command salaries well above local medians. The presence of high-wage workers increases local consumption, supporting retail and service employment. Moreover, facilities require ongoing maintenance, cleaning, security, and other services that may be sourced locally.

Indirect and induced effects. Beyond direct employment, data centers may generate broader economic spillovers. New establishments may form to serve data center operations—IT services firms, equipment suppliers, specialized contractors. Existing firms may expand to meet increased demand. The presence of a major technology investment may signal the quality of local infrastructure, attracting additional investment. Conversely, data centers may crowd out other activities by competing for land, labor, and especially electricity.

Electricity market effects. Data centers are exceptionally electricity-intensive. A single hyperscale facility consuming 100+ megawatts continuously draws approximately 876,000 megawatt-hours annually—equivalent to the demand of roughly 80,000 homes. This concentrated load may strain local generation and transmission capacity, particularly in regions with limited reserve margins. If utilities must procure additional power or invest in transmission upgrades, these costs may be passed through to all ratepayers in the form of higher electricity prices. The magnitude of this effect depends on local grid conditions, the pace of data center growth, and regulatory frameworks governing cost allocation.

The net local impact of data center entry depends on how these channels interact and

which effects dominate. Construction creates immediate but temporary employment gains. Operations create persistent but modest direct employment. Spillovers may amplify these effects if local conditions support multiplier mechanisms, or may be limited if data centers operate as relatively isolated facilities with few local linkages.

2.4 Hypotheses

Drawing on the theoretical framework and existing literature, we develop five hypotheses about the local economic effects of data center entry.

Positive Local Effects. The million-dollar plants literature consistently finds that large facility openings generate positive local spillovers. Greenstone et al. (2010) document productivity gains for incumbent plants; Moretti (2010) estimates employment multipliers of 1.6 for tradable-sector jobs. If data centers function similarly to other large investments, we expect positive effects on local economic activity.

Several mechanisms support this prediction. Construction employment directly increases local payrolls during the building phase. Operations employment, while modest in scale, is high-wage and may generate consumption spillovers. The presence of a major facility may attract complementary investment from suppliers and service providers. Local governments may experience increased tax revenues that support public services and infrastructure.

However, countervailing forces may limit these effects. Data centers employ far fewer permanent workers than comparably-sized manufacturing plants—a $500 million data center might employ 50 workers, while a $500 million auto plant might employ 2,000. If employment multipliers scale with direct job creation, data center spillovers may be correspondingly smaller. Moreover, data centers' limited local supply chains—most equipment is manufactured elsewhere and shipped in—may reduce backward linkages that drive multiplier effects in manufacturing. The net effect is an empirical question.

Hypothesis 1: Data center entry increases local employment, wages, and business establishments.

Operator Heterogeneity. Not all data centers are alike. Facilities operated by major technology companies—Amazon Web Services, Google, Microsoft Azure, Meta, Apple, and Oracle—differ systematically from third-party colocation and enterprise data centers. Big Tech facilities are typically hyperscale operations designed for cloud computing and AI workloads, with capacities of 50–500+ megawatts. Third-party facilities tend to be smaller (5–50 megawatts) and serve diverse tenants with varying requirements.

These differences have implications for local economic effects. Hyperscale facilities

require more extensive construction, with longer build-out periods and larger workforces. Their enormous electricity requirements may necessitate dedicated transmission infrastructure, creating additional construction employment. Big Tech companies often make long-term commitments to host communities, including direct investments in local education, workforce development, and community amenities. These commitments may generate goodwill and complementary investment beyond what third-party operators provide.

Hypothesis 2: Big Tech facilities generate larger local effects than third-party data centers.

Metro Concentration. The magnitude of local multipliers depends critically on the economic environment into which investment flows. Moretti (2012) emphasizes that agglomeration economies—thick labor markets, specialized suppliers, knowledge spillovers—amplify the effects of new investment in already-dense locations. Conversely, investments in thin markets may generate fewer spillovers because the complementary inputs required for multiplier effects are absent.

Metropolitan areas offer several advantages for realizing data center spillovers. Deep labor markets allow firms to hire specialized workers without costly relocation. Existing supplier networks can scale up to serve new customers. Dense service sectors absorb increased consumer spending from high-wage data center workers. These conditions create fertile ground for multiplier effects.

Non-metropolitan areas face structural disadvantages. Thin labor markets may require importing workers from elsewhere, limiting local employment gains. Absent supplier networks, data centers may rely entirely on non-local procurement. Limited local services mean consumption spillovers may leak to larger nearby cities. While data centers in rural areas receive considerable attention—they are often more visible and generate more local controversy—their economic effects may be correspondingly limited.

Hypothesis 3: Local economic effects are concentrated in metropolitan areas, with limited impacts in non-metro counties.

Entry Timing. The data center industry has evolved dramatically over our study period. Early facilities (pre-2010) were often pioneering investments in locations with limited existing digital infrastructure. A new data center in 2005 may have represented a transformative shock to the local economy, bringing not only direct employment but also signaling the region's viability for technology investment.

Recent entrants face a different landscape. Data center clusters have emerged in Northern Virginia, Dallas-Fort Worth, Phoenix, and other markets, with dozens of facilities operating in close proximity. New facilities in these established markets represent a smaller marginal shock. Moreover, recent entrants can draw on existing pools of specialized labor

and established supplier relationships, reducing their need to develop local capacity.

Hypothesis 4: Earlier data center entrants experienced larger local effects than recent entrants.

Electricity Prices. Data centers rank among the most electricity-intensive facilities in the economy. A hyperscale data center consuming 100 megawatts continuously draws approximately 876,000 megawatt-hours annually—more than many small cities. This concentrated demand has drawn increasing attention from policymakers, utilities, and ratepayer advocates concerned about cost allocation and grid reliability.

Economic theory predicts that large demand shocks should increase prices through several channels. In the short run, utilities may need to procure additional power from more expensive marginal generators. If demand exceeds local generation capacity, power must be imported over transmission lines, incurring congestion costs. In the longer run, utilities may need to invest in new generation or transmission capacity, with costs recovered through rate increases approved by public utility commissions.

Hypothesis 5: Data center entry increases local electricity prices.

3 Data and Empirical Approach

3.1 Data Sources

We combine facility-level data on U.S. data centers with county-level measures of economic activity and electricity prices.

3.1.1 Data Center Data

Our primary source for data center information is the Aterio dataset. Aterio is a commercial data provider that compiles the most comprehensive facility-level coverage of U.S. data centers, aggregating information from utility filings, permit records, satellite imagery, and industry sources. For each facility, the dataset reports: exact street-level location (latitude, longitude, county FIPS code); activation date and current operational status (active, announced, under construction, canceled); facility type and operator name; utility connections; and estimated capacity measured in megawatts (MW).

This granularity allows us to identify not only where data centers are located but also their scale and timing of entry. We define treatment based on the activation of the first data center in a county, which provides a clean adoption date for our staggered difference-in-differences design. To test our hypotheses, we construct several key classification variables:

- **Operator type:** Whether the first facility is operated by a "Big Tech" company (Amazon, Google, Microsoft, Meta, Apple, or Oracle) versus other operators.
- **Metro status:** Whether the county is classified as metropolitan or non-metropolitan based on Office of Management and Budget (OMB) core-based statistical area delineations.
- **Entry cohort:** The timing of first data center activation, classified as pre-2010, 2010–2019, or post-2020.

We also construct additional heterogeneity variables for supplementary analyses: data center size measured by MW capacity (median split among treated counties); rural-urban continuum codes (RUCC) from the USDA Economic Research Service, which classify counties on a 1–9 scale based on metropolitan status and adjacency to metro areas; and county population measures including total population and population density (persons per square kilometer) from the Census Bureau's Population Estimates Program.

3.1.2 Economic Outcomes

Our primary outcome data come from the Quarterly Census of Employment and Wages (QCEW), which provides county-level information on employment, total wages, and establishment counts at quarterly frequency. The QCEW covers virtually all employers subject to state unemployment insurance laws, representing approximately 95% of U.S. jobs. We focus on total private-sector employment and wages, using log transformations to facilitate proportional interpretation of treatment effects.

We supplement the QCEW with additional county-level outcomes: median household income from the Census Bureau's Small Area Income and Poverty Estimates (SAIPE); unemployment rates from the Bureau of Labor Statistics Local Area Unemployment Statistics (LAUS); and building permits from the Census Bureau's Building Permits Survey. These outcomes provide complementary perspectives on local economic conditions.

3.1.3 Electricity Data

For electricity outcomes, we use data from the Energy Information Administration (EIA) Form 861M, which reports monthly retail electricity prices by utility. A key challenge is that utility service territories do not align with county boundaries, creating two distinct measurement issues.

First, many counties are served by multiple utilities. When we aggregate to the county level, we compute sales-weighted average prices across all serving utilities using EIA service

territory files. This "blending" of prices adds measurement noise but does not systematically bias our estimates.

Second, and more problematic for causal inference, many utilities serve multiple counties simultaneously. When a data center enters one county and increases electricity demand, the utility's response—whether procuring additional power or raising rates—may affect all counties in its service territory similarly. This creates "price spillover": control counties that share utilities with treated counties experience the same price changes, attenuating the treatment-control comparison toward zero.

To quantify this identification challenge, we compute a "spillover exposure" measure for each county: the number of unique other counties that share at least one utility provider. The median county shares utilities with 63 other counties; only 23 counties (0.7% of all U.S. counties) face zero spillover exposure. This pervasive interconnection means that conventional county-level comparisons will understate the true effect of data center entry on electricity prices.

3.2 Sample Construction

We merge the Aterio facility data with economic outcomes by county FIPS code, creating a county-quarter panel spanning 2000Q2 through 2024Q4. Our sample begins in 2000Q2 to ensure adequate pre-treatment observations for early-adopting counties while avoiding periods with sparse data center activity.

Treatment is defined as the quarter of first data center activation in a county. Counties that never host a data center serve as the primary control group ("never-treated"). We also report results using not-yet-treated counties as controls for robustness (see Appendix B).

Table 1 presents summary statistics for treated and control counties. Our sample includes 447 treated counties and 2,846 control counties over 99 quarters (2000Q2–2024Q4). Treated counties are substantially larger than controls: mean log employment is 11.1 for treated counties versus 8.8 for controls. This difference reflects the economic geography of data center location—facilities cluster in larger, more developed areas with access to fiber networks, reliable power, and skilled labor pools. Rather than invalidating our analysis, this pattern motivates our heterogeneity analysis by metro status: we examine whether effects differ between metropolitan and non-metropolitan treated counties.

Approximately 78% of treated counties are metropolitan, compared to only 2% of controls. Among treated counties, median data center capacity is 12.5 MW, with a mean of 107 MW—indicating substantial right-skew driven by a small number of very large hyperscale facilities. Approximately 15% of first entries are Big Tech facilities.

3.3 Empirical Approach

3.3.1 Callaway-Sant'Anna Estimator

We estimate treatment effects using the Callaway and Sant'Anna (2021) difference-in-differences estimator for staggered adoption. This approach addresses well-documented problems with traditional two-way fixed effects (TWFE) estimators, which can produce biased estimates when treatment effects are heterogeneous across cohorts or over time. Specifically, TWFE can assign negative weights to some treated observations, causing already-treated units to serve as implicit controls for later-treated units—a comparison that conflates treatment dynamics with treatment effects (Goodman-Bacon, 2021; Sun and Abraham, 2021).

Let G_c denote the quarter in which county c first activates a data center, with $G_c = \infty$ for never-treated counties. The Callaway-Sant'Anna estimator computes group-time average treatment effects:

$$ATT(g,t) = E\left[Y_{ct}(1) - Y_{ct}(0) \mid G_c = g\right]$$

for each cohort g (counties treated in the same quarter) and time period $t \geq g$. Crucially, these effects are identified by comparing treated counties only to valid controls—never-treated or not-yet-treated counties—rather than to already-treated counties that may have different treatment dynamics.

The overall average treatment effect aggregates across cohorts and time periods:

$$ATT = \sum_{g} \sum_{t \geq g} w_{g,t} \cdot ATT(g,t)$$

where weights $w_{g,t}$ reflect the share of observations in each group-time cell. We report results using never-treated counties as controls in our main specifications, with not-yet-treated controls as robustness checks in Appendix B.

All specifications include county-level covariates to improve precision: log population, log land area, and demographic composition (percent white, percent aged 20–39, percent aged 40–64, percent male). These are slow-moving characteristics that are plausibly unaffected by data center entry, satisfying the requirement that covariates in staggered DiD not be influenced by treatment.

3.3.2 Event Study Specification

To examine treatment effect dynamics, we estimate event study specifications that trace effects relative to the activation quarter. Define event time as $e = t - G_c$, the number of

quarters relative to treatment. The event study reports:

$$ATT(e) = \sum_{g} w_g \cdot ATT(g, g+e)$$

which averages effects across cohorts at each event time. Estimates for $e < 0$ provide a test of the parallel trends assumption: if treated and control counties were evolving similarly before treatment, pre-treatment effects should be close to zero. Estimates for $e \geq 0$ trace the post-treatment trajectory of outcomes.

We report event studies spanning 12 quarters before through 12 quarters after treatment, with the quarter immediately before activation ($e = -1$) as the reference period. Certain counties are automatically dropped from the estimation to ensure balanced panels.

3.3.3 Heterogeneity Analysis

To test Hypotheses 2–4, we stratify our sample along three dimensions and estimate separate treatment effects within each stratum:

Operator type: We classify counties based on whether the first data center was operated by a Big Tech company (Amazon, Google, Microsoft, Meta, Apple, or Oracle) or another operator. Big Tech facilities represent approximately 15% of first entries in our sample.

Metro status: We classify counties as metropolitan or non-metropolitan using OMB delineations based on core-based statistical areas (CBSAs). Metropolitan counties contain or are economically integrated with urbanized areas of 50,000+ population. Approximately 78% of treated counties are metropolitan.

Entry cohort: We classify counties into three cohorts based on the year of first data center activation: early entrants (pre-2010), middle entrants (2010–2019), and recent entrants (post-2020). This captures the evolution of the data center industry from early enterprise facilities through the cloud computing boom to the current AI-driven expansion.

For each stratification, we estimate the Callaway-Sant'Anna model separately within each group and compare the resulting treatment effects. Additional heterogeneity analyses by data center size, county population, and RUCC category are presented in Appendix C.

3.3.4 Electricity Identification Strategy

Testing Hypothesis 5 requires addressing the geographic mismatch between utility service territories and county boundaries. As discussed above, when utilities serve multiple counties, treatment effects diffuse across jurisdictional boundaries, attenuating the estimated effect

toward zero.

Our strategy exploits variation in spillover exposure across counties. Counties with low spillover exposure—those sharing utilities with fewer other counties—provide cleaner identification because price changes more directly reflect local demand conditions. Counties with high spillover exposure face greater measurement challenges, as any treatment effect is diluted across many counties.

We operationalize this by computing the median spillover exposure among treated counties in our restricted sample and dividing counties into below-median (low spillover) and above-median (high spillover) groups. We then estimate treatment effects separately for each group. If price spillover is indeed attenuating our estimates, we expect to find:

- Positive, statistically significant effects in low-spillover counties, where identification is cleaner.
- Attenuated or null effects in high-spillover counties, where measurement challenges dominate.

This gradient across spillover categories serves as a diagnostic: if effects appear only where spillover is limited, we can be more confident that the null findings in high-spillover areas reflect measurement challenges rather than a true absence of price effects.

A further challenge is the inherent magnitude of the demand shock. The median first-entry data center in our sample has a nameplate capacity of 10 MW. At realistic utilization rates—accounting for Power Usage Effectiveness (PUE) overhead and the gap between nameplate and actual power draw, effective consumption is roughly 50% of nameplate—this translates to approximately 44 GWh of annual electricity consumption. The median utility serving a treated county at the time of first entry sells approximately 3,600 GWh per year. The first-entry demand shock therefore represents roughly 1.2% of the typical utility's total load. Figure 4 illustrates the distribution of first-entry capacity and utility size. For nearly half of treated counties, the first data center's demand is below 1% of the serving utility's annual sales. This means that even with clean identification at the county level, the price signal from a single data center entry is inherently modest, and detectable effects require either large facilities relative to local utility scale or the cumulative impact of subsequent entries over time.

4 Results

4.1 Summary Statistics and Correlations

Table 1 presents summary statistics for our analysis sample. Panel A reports economic outcomes at the county-quarter level. Treated counties are substantially larger than controls across all measures: mean log employment is 11.1 versus 8.8, mean log wages is 20.3 versus 17.9, and mean log establishments is 8.3 versus 6.4. These differences of approximately 2–2.5 log points correspond to treated counties being roughly 7–10 times larger than controls in levels. Treated counties also have higher median household incomes ($53,574 vs. $43,284) and slightly lower unemployment rates (5.5% vs. 5.9%).

Panel B reports electricity prices. Treated counties face somewhat higher prices across all sectors: residential prices average $117/MWh versus $110/MWh for controls, with similar patterns for commercial and industrial rates. These differences may reflect both selection (data centers locate in areas with certain grid characteristics) and treatment effects.

Panel C describes data center characteristics among treated counties. The distribution of facility size is highly right-skewed: median capacity is 12.5 MW, but mean capacity is 107 MW, driven by a small number of hyperscale facilities. Approximately 15% of first entries are Big Tech facilities. Entry years span from the 1990s through 2024, though our empirical analysis uses data from 2000Q2 onward to ensure adequate pre-treatment observations.

Panel D reports county characteristics. Treated counties have substantially larger populations (mean 437,000 vs. 52,000) and higher population density (996 vs. 155 per km^2). Treated counties are also more likely to be metropolitan (78% versus 2% of controls), reflecting the tendency of data centers to locate in areas with existing infrastructure. This selection pattern motivates our heterogeneity analysis by metro status.

Table 2 presents correlations among heterogeneity characteristics for treated counties. Entry cohort is positively correlated with Big Tech status (0.29), reflecting the recent growth of hyperscale cloud facilities. Metro status (as captured by RUCC category) is positively correlated with data center size (0.16) and population density (0.68). These correlations inform interpretation of our heterogeneity results—for example, Big Tech effects may partly reflect the characteristics of more recent entry cohorts.

4.2 Baseline Results

Table 3 presents our main estimates of the effect of data center entry on local economic outcomes using never-treated counties as controls. We find positive and statistically significant effects across most outcomes.

Data center entry increases log employment by 0.035 (SE = 0.010), corresponding to approximately 3.5% higher employment in treated counties relative to controls. This effect is statistically significant at the 1% level. Log total wages increase by 0.050 (SE = 0.012), or approximately 5.0%, also significant at the 1% level. The larger wage effect relative to the employment effect suggests that data center entry may raise average wages, consistent with the creation of high-wage jobs.

Log establishments increase by 0.047 (SE = 0.011), indicating 4.7% more business establishments in treated counties. Log median household income increases by 0.019 (SE = 0.005), a more modest 1.9% gain that is nonetheless statistically significant. The unemployment rate shows a small and statistically insignificant decline of 0.03 percentage points. Log building permits increase substantially by 0.161 (SE = 0.046), or approximately 16%, consistent with construction activity associated with data center development.

Figures 5–7 present event study estimates for all six outcomes. The pre-treatment coefficients are generally close to zero and statistically insignificant, supporting the parallel trends assumption. Treatment effects emerge around the activation quarter and persist through the post-treatment period, with some evidence of gradual increase over time.

These baseline results provide support for Hypothesis 1: data center entry generates positive local economic effects. The magnitudes are economically meaningful, though somewhat smaller than effects documented in the million-dollar plants literature for large manufacturing facilities.

4.3 Heterogeneity by Operator Type

Table 4 presents results stratified by whether the first data center in a county was operated by a Big Tech company (Amazon, Google, Microsoft, Meta, Apple, or Oracle). Consistent with Hypothesis 2, Big Tech facilities generate larger effects than non-Big Tech facilities.

For log employment, the effect in Big Tech counties is 0.040 (SE = 0.024), compared to 0.036 (SE = 0.012) in non-Big Tech counties. For wages, Big Tech counties show a 0.072 effect (SE = 0.026) versus 0.047 (SE = 0.013) for non-Big Tech. For establishments, the pattern is similar: 0.049 (SE = 0.020) for Big Tech versus 0.048 (SE = 0.013) for non-Big Tech.

Figures 8–10 display event study estimates separately for Big Tech and non-Big Tech counties across all six outcomes. Both groups show parallel pre-trends, but the post-treatment divergence is larger for Big Tech facilities, particularly for wages. These results suggest that the scale and characteristics of hyperscale Big Tech facilities translate into meaningfully larger local economic impacts.

4.4 Heterogeneity by Metro Status

Table 5 presents results stratified by metropolitan status. The results strongly support Hypothesis 3: effects are concentrated in metropolitan areas while non-metro counties show little impact.

For log employment, the metro effect is 0.041 (SE = 0.010), statistically significant at the 1% level. The non-metro effect is essentially zero: 0.001 (SE = 0.032). For wages, metro counties show a 0.055 effect (SE = 0.012) while non-metro counties show 0.017 (SE = 0.037), not statistically significant. A similar pattern appears for establishments: metro counties experience a 0.050 increase (SE = 0.013) while non-metro counties show 0.034 (SE = 0.015).

Interestingly, non-metro counties show a significant decline in unemployment rates (-0.73 percentage points, SE = 0.33) that is not apparent in metro counties. This may reflect local labor market tightening in thin rural markets even without substantial employment gains in our measured sectors.

Figures 11–13 display the stark contrast between metro and non-metro event studies across all six outcomes. Metro counties show clear positive effects emerging after treatment, while non-metro estimates are noisy and centered near zero. These results highlight the importance of agglomeration economies for realizing the local benefits of data center investment.

4.5 Heterogeneity by Entry Cohort

Table 6 presents results stratified by the timing of first data center entry. Consistent with Hypothesis 4, earlier entrants show stronger effects than recent entrants, though the pattern varies across outcomes.

For log employment, pre-2010 entrants show an effect of 0.033 (SE = 0.014), 2010–2019 entrants show 0.050 (SE = 0.015), and post-2020 entrants show 0.007 (SE = 0.007). For wages, the pattern is similar: effects of 0.060, 0.054, and 0.018 for the three cohorts respectively. The largest effects appear in the middle cohort (2010–2019), which coincides with the cloud computing boom and may represent the sweet spot of market maturity.

Figures 14–16 display event study estimates by entry cohort for all six outcomes. Pre-2010 and 2010–2019 entrants show clearer positive trajectories, while post-2020 entrants show more muted effects. These results suggest either that the local economic benefits of data center entry may be declining as the industry matures and established clusters reduce the marginal impact of new facilities, or that effects from more recent entry cohorts have not yet had sufficient time to materialize fully given their shorter post-treatment observation

windows. Note that this cohort timing analysis compares counties that received their *first* data center in different eras; we turn next to the question of whether *additional* data center entries in already-treated counties generate incremental benefits.

4.6 Additional Entrants Analysis

Our main analysis examines the effect of a county's first data center entry. However, many counties experience multiple data center entries over time, raising the question of whether additional entries amplify or attenuate local economic effects. To investigate this intensive margin, we classify treated counties based on whether they received additional data centers within five years of their first entry.

Figure 3 illustrates our classification approach. We define *Isolated Entry* counties as those where the first data center remains the only entry within a five-year window (57.3% of treated counties, N=118). *Clustered Entry* counties receive one or more additional data centers within five years of the first entry (42.7%, N=88). We select the five-year window to balance two considerations: shorter windows capture too few additional entries for meaningful variation, while longer windows may conflate clustering effects with other time-varying factors.

Table 7 presents the results. Clustered entry counties exhibit consistently larger effects than isolated entry counties across all core economic outcomes. For employment, clustered counties show a 4.3% increase compared to 3.6% for isolated counties. Similar patterns appear for wages (7.7% vs. 6.0%) and establishments (5.9% vs. 5.1%). Household income effects are also larger for clustered counties (3.5% vs. 2.0%), as are building permits (25.1% vs. 20.9%). Unemployment rate effects are noisy and statistically insignificant for both groups.

These results suggest that data center clustering is associated with larger local economic benefits. However, we interpret this finding cautiously. We cannot distinguish whether the larger effects in clustered counties reflect additive contributions from multiple facilities, complementarities between co-located data centers, or selection—locations that attract multiple entries may possess unobserved characteristics that amplify economic spillovers. Nonetheless, the pattern is consistent with the hypothesis that data center agglomeration generates cumulative local benefits.

4.7 Electricity Price Effects

Finally, we turn to Hypothesis 5 regarding electricity prices. As discussed in Section 3, identification is complicated by the geographic structure of utility service territories, which

span multiple counties and create spillovers across jurisdictional boundaries.

Figure 17a presents baseline event study estimates for log total electricity price. The estimates are noisy with wide confidence intervals, and we do not detect a clear treatment effect in the full sample. This null result likely reflects both the modest magnitude of the first-entry demand shock—roughly 1% of the typical utility's annual load, as documented in Section 3—and the diffusion of price effects across utility service territories spanning multiple counties.

To address these challenges, we stratify counties by spillover exposure—the number of other counties sharing utility providers. Table 8 presents results for low-spillover (below median, $\leq$64 shared counties) and high-spillover (above median, >64 shared counties) groups.

In low-spillover counties, where identification is cleaner, we find positive effects on electricity prices. Total price increases by 0.050 (SE = 0.023), statistically significant at the 5% level. Industrial prices show the largest effect (0.066, SE = 0.028), while residential and commercial prices show smaller effects (0.036 and 0.036 respectively).

In high-spillover counties, effects are attenuated and in some cases negative, though not consistently statistically significant. This pattern is consistent with our identification strategy: when treatment effects diffuse across many counties through shared utilities, the comparison between treated and control counties fails to detect localized effects.

Figure 17b displays event study estimates separately for low and high spillover counties. Low-spillover counties show a clearer positive trajectory after treatment, while high-spillover estimates are noisy.

These results provide qualified support for Hypothesis 5. Data center entry appears to raise local electricity prices, but the effect is only detectable in areas with cleaner identification. The prevalence of price spillover across utility territories means that the full magnitude of electricity cost increases may be larger than our estimates suggest, as costs are distributed across broader geographic areas.

4.8 Robustness Checks and Additional Analyses

We conduct several robustness checks and additional analyses in the appendices. Appendix A presents results using only geographically adjacent counties as controls, providing a more localized comparison. Appendix B reports estimates using not-yet-treated counties as controls rather than never-treated counties. Appendix C presents additional heterogeneity analyses by data center size, county population, and RUCC category. These supplementary analyses generally support the robustness of our main findings.

5 Discussion

Our results provide systematic evidence on the local economic effects of data center entry. We find positive effects on employment, wages, and establishments, with heterogeneity along several dimensions. Here we discuss the interpretation, implications, and limitations of these findings.

5.1 Magnitude and Mechanisms

The baseline employment effect of 3.5% is economically meaningful but smaller than effects documented for large manufacturing plants. Greenstone et al. (2010) found productivity gains of 12% for incumbent plants following million-dollar plant openings, while Moretti (2010) estimated employment multipliers of 1.6 for tradable-sector jobs. Our smaller effects likely reflect the fundamental difference between data centers and manufacturing: a $500 million data center employs perhaps 50 permanent workers, while a comparably-sized manufacturing plant might employ 2,000.

What mechanisms drive the effects we observe? Construction employment is one channel: the 16% increase in building permits suggests substantial construction activity around data center entry. However, construction effects should be temporary, while our event studies show persistent effects over multiple years. This persistence suggests that ongoing operations employment and induced effects on local services also contribute.

The larger effect on total wages (5.0%) relative to employment (3.5%) implies an increase in average wages. This is consistent with data centers creating high-wage positions that raise the local wage bill disproportionately to headcount. The 4.7% increase in establishments suggests new firm formation or entry, possibly including IT service providers, contractors, and businesses serving data center workers.

The concentration of effects in metropolitan areas highlights the importance of complementary inputs. Data centers in metro areas can draw on existing labor pools, supplier networks, and service infrastructure. Rural data centers, lacking these complementary inputs, generate fewer spillovers even when they represent substantial investments relative to the local economy. This finding aligns with Moretti (2012)'s emphasis on agglomeration as a precondition for multiplier effects.

The declining effects among later first-entry cohorts suggest diminishing returns to cluster development. Early first entrants generate outsized benefits by establishing foundational infrastructure, workforce capabilities, and supply chain networks that create multiplier effects across local industries. The weaker effects observed for more recent first entrants may

reflect shorter post-treatment observation windows, as these facilities have had less time to generate cumulative economic impacts.

Our analysis of entry clustering reveals that counties experiencing multiple data center entries within five years of the initial entry show larger treatment effects than counties with isolated first entries. Clustered counties exhibit employment gains of 4.3% compared to 3.6% for isolated counties, with similar patterns for wages, establishments, and building permits. This pattern suggests that data center clustering generates cumulative benefits, though we cannot distinguish whether this reflects additive effects, complementarities between facilities, or selection on unobserved location characteristics.

Taken together, these heterogeneity patterns suggest that location characteristics serve as a precondition for local economic spillovers from data center entry. The limited effects observed in non-metropolitan counties indicate that capital investment alone is insufficient to generate local gains in the absence of complementary labor markets and business density. Conditional on a favorable local environment, facility-level characteristics such as operator type, entry timing, and clustering then modulate the magnitude of those spillovers.

5.2 Policy Implications

Our findings inform several ongoing policy debates. The positive economic effects documented here—increases in employment, wages, and business activity—provide some justification for state and local incentives used to attract data centers. These represent real benefits to local communities. However, the magnitude of these effects, while statistically significant, is modest compared to the scale of subsidies often provided. A careful cost-benefit analysis would need to weigh the economic gains against the fiscal cost of tax abatements, infrastructure investments, and other incentives.

Several caveats temper these positive conclusions. First, the concentration of benefits in metropolitan areas complicates rural economic development strategies. Several states have targeted data centers as engines of rural revitalization, attracted by their land requirements and relatively modest labor needs. Our results suggest this strategy may be misguided: non-metro counties experience few spillover benefits from data center entry. Policymakers hoping to use data centers for rural development should temper expectations and consider whether alternative investments might generate larger local impacts.

Second, our electricity results highlight distributional concerns that deserve policy attention. If data center demand raises electricity prices for all customers in a utility's service territory, the burden falls on residential ratepayers and certain commercial customers who do not directly benefit from data center employment. This hidden cost may offset some of the

visible economic benefits, particularly in regions where data center growth is rapid. States might consider requiring data centers to bear a greater share of infrastructure costs, either through direct contributions or through rate structures that allocate costs more precisely.

Third, our results on entry timing and clustering jointly suggest that incentive policies should consider both when and how many data centers enter a locality. The declining effects among recent first-entry cohorts indicate that pioneering investments may generate larger benefits than later arrivals to established markets. However, our clustering analysis reveals that counties receiving additional entries within five years of the first show amplified effects relative to isolated entries. Together, these findings suggest that while the marginal benefit of attracting a first data center may be declining over time, locations that have already secured an initial facility may benefit from policies that encourage subsequent entries—potentially offsetting some of the diminishing returns to later market entry.

5.3 Limitations

Several limitations of our analysis warrant discussion. First, we cannot directly observe the occupation or skill composition of jobs affected by data center entry. While we hypothesize that high-wage operations jobs drive consumption spillovers, we cannot test this mechanism with our county-level data. More granular data on occupational employment would allow us to trace the transmission channels more precisely.

Second, our electricity analysis faces persistent identification challenges despite our spillover-stratified approach. Even among low-spillover counties, some cross-county contamination likely remains. The true magnitude of electricity price effects may be larger than our estimates suggest, with costs distributed across broader geographic areas than individual counties. Moreover, because our design captures the effect of a single first entry—which for the median county represents roughly 1% of the serving utility's total annual sales—the cumulative effect of subsequent entries and capacity expansion over time may be substantially larger than our first-entry estimates indicate.

Third, we focus on relatively short-run effects in our event study graphs spanning three years after data center activation. The long-run consequences of data center investment—including potential cluster development, infrastructure strain, or changes in local industrial composition—may differ from these short-run patterns. Extended panel data as the current wave of AI-driven investment matures will allow analysis of longer-term dynamics. Relatedly, while our additional entrants analysis provides suggestive evidence that multiple data center entries amplify local effects, we compare counties with different entry intensities rather than estimating the marginal effect of each additional facility.

Fourth, our identification relies on comparing treated counties to never-treated or not-yet-treated controls. If data center location decisions respond to anticipated local economic trajectories, this comparison may be confounded. The flat pre-trends in our event studies provide some reassurance, but we cannot fully rule out selection on unobservable growth potential.

6 Conclusion

The explosive growth of artificial intelligence has triggered an unprecedented wave of data center investment, with capital expenditure rivaling the great infrastructure buildouts of American history. This paper provides the first comprehensive analysis of how this investment wave affects local economies.

Using detailed facility-level data merged with county-level outcomes and employing modern difference-in-differences methods, we find that data center activation increases local employment by 3.5%, wages by 5.0%, and establishments by 4.7%. These effects are economically meaningful and statistically robust, providing evidence that data centers do generate local economic benefits despite their relatively modest direct employment.

However, the distribution of these benefits is highly uneven. Metropolitan counties capture the gains; non-metro counties experience negligible effects. Big Tech hyperscale facilities generate larger impacts than smaller third-party data centers. Counties with early first-entry timing experienced stronger effects than those with recent first-entry timing, and counties that received additional data centers within five years of first entry showed larger impacts than those with isolated entries. For communities hoping to benefit from the AI boom, these patterns suggest that the characteristics of the host community—not just the facility itself—shape whether investment translates into local economic gains.

We also document a hidden cost: data center entry appears to raise electricity prices in areas where utility structure allows clean identification. As AI computing demand continues to accelerate, the strain on electricity infrastructure—and the resulting cost burden on ratepayers—deserves careful policy attention.

Looking forward, several questions merit further research. How do data center effects evolve over longer time horizons? Do clusters develop internal agglomeration economies that compound initial gains? How do communities balance economic benefits against quality-of-life concerns that have prompted local opposition? And as AI models grow ever larger and more power-hungry, how will electricity grids adapt to unprecedented demand growth?

The data center industry stands at an inflection point. The investments being made today will shape local economies for decades to come. Understanding their effects—both

benefits and costs—is essential for communities, policymakers, and investors navigating this transformation. This study provides an empirical foundation for these deliberations, offering the first systematic evidence on how data center entry reshapes local economies and highlighting the heterogeneous nature of these effects across different contexts.

References

Busso, M., Gregory, J., and Kline, P. (2013). Assessing the incidence and efficiency of a prominent place based policy. *American Economic Review*, 103(2):897–947.

Callaway, B. and Sant'Anna, P. H. (2021). Difference-in-differences with multiple time periods. *Journal of Econometrics*, 225(2):200–230.

Fang, T. and Greenstein, S. (2025). Where the cloud rests: The economic geography of data centers. *Strategy Science*, 10(4):404–420.

Goodman-Bacon, A. (2021). Difference-in-differences with variation in treatment timing. *Journal of Econometrics*, 225(2):254–277.

Greenstone, M., Hornbeck, R., and Moretti, E. (2010). Identifying agglomeration spillovers: Evidence from winners and losers of large plant openings. *Journal of Political Economy*, 118(3):536–598.

Greenstone, M. and Moretti, E. (2004). Bidding for industrial plants: Does winning a million dollar plant increase welfare?

Kline, P. and Moretti, E. (2014). Local economic development, agglomeration economies, and the big push: 100 years of evidence from the tennessee valley authority. *The Quarterly Journal of Economics*, 129(1):275–331.

Masanet, E., Shehabi, A., Lei, N., Smith, S., and Koomey, J. (2020). Recalibrating global data center energy-use estimates. *Science*, 367(6481):984–986.

Moretti, E. (2010). Local multipliers. *American Economic Revieww: Papers & Proceedings*, 100(2):373–377.

Moretti, E. (2012). *The new geography of jobs*. Houghton Mifflin Harcourt.

Mullin, J. (2023). Features: Virginia's data centers and economic development. *Econ Focus*, 23(2Q):14–17.

Park, J., Han, K., and Lee, B. (2023). Green cloud? An empirical analysis of cloud computing and energy efficiency. *Management Science*, 69(3):1639–1664.

Pham, N. (2017). Data centers: Jobs and opportunities in communities nationwide. *Available at SSRN 2998644*.

Shehabi, A., Smith, S. J., Masanet, E., and Koomey, J. (2018). Data center growth in the United States: decoupling the demand for services from electricity use. *Environmental Research Letters*, 13(12):124030.

Siddik, M. A. B., Shehabi, A., and Marston, L. (2021). The environmental footprint of data centers in the united states. *Environmental Research Letters*, 16(6):064017.

Slattery, C. (2025). Bidding for firms: Subsidy competition in the united states. *Journal of Political Economy*, 133(8):000–000.

Slattery, C. and Zidar, O. (2020). Evaluating state and local business incentives. *Journal of Economic Perspectives*, 34(2):90–118.

Sun, L. and Abraham, S. (2021). Estimating dynamic treatment effects in event studies with heterogeneous treatment effects. *Journal of Econometrics*, 225(2):175–199.

Tables

Table 1: Summary Statistics

	Treated (N=447 counties)			Control (N=2,846 counties)			
Variable	Mean	SD	Median	Mean	SD	Median	Diff (T-C)
Panel A: Economic Outcomes (Quarterly, 2000–2024)							
Log Employment	11.1	1.67	11.2	8.84	1.39	8.79	2.22***
Log Total Wages	20.3	1.84	20.4	17.9	1.56	17.8	2.44***
Log Establishments	8.33	1.56	8.36	6.37	1.23	6.26	1.96***
Median HH Income ($)	53,574	18,494	49,970	43,284	13,673	40,814	10,290***
Unemployment Rate (%)	5.5	2.3	5.0	5.9	3.0	5.2	−0.4***
Building Permits	1,657	3,078	621	205	588	42	1,452***
Panel B: Electricity Prices (Quarterly, 2010–2024)							
Residential ($/MWh)	117.3	41.6	110.9	109.7	39.6	104.6	7.64***
Commercial ($/MWh)	97.9	35.0	91.7	92.5	35.0	87.8	5.39***
Industrial ($/MWh)	71.8	34.0	63.8	65.9	39.4	60.9	5.95***
Total ($/MWh)	98.9	39.0	91.4	90.5	36.4	85.2	8.41***
Panel C: Data Center Characteristics (Treated Only)							
Capacity (MW)	106.7	388.7	12.5	—	—	—	—
Big Tech (%)	15.4	—	—	—	—	—	—
Entry Year	2015	12	2016	—	—	—	—
Panel D: County Characteristics							
Population (000s)	436.9	758.4	200.4	52.4	117.2	21.3	384.6***
Pop. Density (per km^2)	995.7	3,700	326.5	154.7	1,133	37.1	841.0***
Metro County (%)	78.1	—	—	1.7	—	—	76.4***

Notes: Sample: 2000Q2–2024Q4 (99 quarters). Economic outcomes from QCEW (log-transformed). Electricity prices from EIA-861M (2010–2024). Diff = Treated mean − Control mean.
* p<0.10, ** p<0.05, *** p<0.01 from Welch's t-test.

Table 2: Correlation Matrix of Heterogeneity Variables (Treated Counties)

	(1)	(2)	(3)	(4)	(5)	(6)
(1) Entry Cohort	1.00					
(2) DC Size (MW)	0.52***	1.00				
(3) Big Tech	0.29***	0.40***	1.00			
(4) RUCC Category	0.36***	0.16***	0.09*	1.00		
(5) High Population	−0.29***	−0.20***	−0.02	−0.34***	1.00	
(6) Log Pop Density	−0.52***	−0.27***	−0.20***	−0.68***	0.59***	1.00

* p<0.10, ** p<0.05, *** p<0.01. N = 447 treated counties.
Entry Cohort: 1 = ≤2009, 2 = 2010–2019, 3 = post-2020.
DC Size: 0 = below median, 1 = at/above median. Big Tech: 0 = no, 1 = yes.
RUCC: 1 = Large Metro, 2 = Medium Metro, 3 = Smaller Areas.

Table 3: CS-DID Baseline Results (Never Treated)

Outcome	ATT	SE	95% CI Lower	95% CI Upper
Log Employment	0.0345***	(0.0100)	0.0149	0.0542
Log Total Wages	0.0496***	(0.0116)	0.0268	0.0723
Log Establishments	0.0467***	(0.0113)	0.0246	0.0689
Log Median HH Income	0.0189***	(0.0049)	0.0093	0.0285
Unemployment Rate	-0.0316	(0.1578)	-0.3409	0.2777
Log Building Permits	0.1613***	(0.0462)	0.0708	0.2519

* p<0.10, ** p<0.05, *** p<0.01
Sample: 2000Q2-Present, 301 treated counties

Table 4: CS-DID Heterogeneity: Big Tech Status

Outcome	Group	ATT	SE	N Treated
Log Employment	Non-Big Tech	0.0358***	(0.0119)	232
Log Employment	Big Tech	0.0399	(0.0242)	69
Log Total Wages	Non-Big Tech	0.0468***	(0.0131)	232
Log Total Wages	Big Tech	0.0715***	(0.0255)	69
Log Establishments	Non-Big Tech	0.0478***	(0.0133)	232
Log Establishments	Big Tech	0.0489**	(0.0200)	69
Log Median HH Income	Non-Big Tech	0.0168***	(0.0054)	232
Log Median HH Income	Big Tech	0.0289***	(0.0112)	69
Unemployment Rate	Non-Big Tech	0.1526	(0.1412)	232
Unemployment Rate	Big Tech	-0.8978**	(0.4300)	69
Log Building Permits	Non-Big Tech	0.1721***	(0.0564)	232
Log Building Permits	Big Tech	0.1377	(0.1132)	69

* p<0.10, ** p<0.05, *** p<0.01

Table 5: CS-DID Heterogeneity: Metro Status

Outcome	Group	ATT	SE	N Treated
Log Employment	Non-Metro	0.0009	(0.0315)	93
Log Employment	Metro	0.0406***	(0.0104)	208
Log Total Wages	Non-Metro	0.0170	(0.0371)	93
Log Total Wages	Metro	0.0549***	(0.0120)	208
Log Establishments	Non-Metro	0.0335**	(0.0149)	93
Log Establishments	Metro	0.0496***	(0.0128)	208
Log Median HH Income	Non-Metro	0.0200*	(0.0114)	93
Log Median HH Income	Metro	0.0186***	(0.0053)	208
Unemployment Rate	Non-Metro	-0.7313**	(0.3333)	93
Unemployment Rate	Metro	0.0927	(0.1671)	208
Log Building Permits	Non-Metro	-0.0256	(0.1056)	93
Log Building Permits	Metro	0.1887***	(0.0531)	208

* p<0.10, ** p<0.05, *** p<0.01

Table 6: CS-DID Heterogeneity: Entry Cohort (3 groups)

Outcome	Group	ATT	SE	N Treated
Log Employment	<=2009	0.0329**	(0.0143)	36
Log Employment	2010-2019	0.0499***	(0.0149)	53
Log Employment	2020+	0.0073	(0.0072)	212
Log Total Wages	<=2009	0.0596***	(0.0171)	36
Log Total Wages	2010-2019	0.0538***	(0.0179)	53
Log Total Wages	2020+	0.0178	(0.0241)	212
Log Establishments	<=2009	0.0544***	(0.0158)	36
Log Establishments	2010-2019	0.0512***	(0.0167)	53
Log Establishments	2020+	-0.0004	(0.0110)	212
Log Median HH Income	<=2009	0.0277***	(0.0073)	36
Log Median HH Income	2010-2019	0.0115*	(0.0062)	53
Log Median HH Income	2020+	0.0080	(0.0103)	212
Unemployment Rate	<=2009	-0.0100	(0.1927)	36
Unemployment Rate	2010-2019	0.0368	(0.2508)	53
Unemployment Rate	2020+	-0.1343	(0.2982)	212
Log Building Permits	<=2009	0.2748***	(0.0801)	36
Log Building Permits	2010-2019	0.1013	(0.0699)	53
Log Building Permits	2020+	0.0427	(0.0532)	212

* p<0.10, ** p<0.05, *** p<0.01

Table 7: CS-DID Heterogeneity: Entry Clustering (Binary Classification)

Outcome	Group	ATT	SE	N Treated
Log Employment	Isolated Entry	0.0357***	(0.0112)	118
Log Employment	Clustered Entry	0.0427***	(0.0128)	88
Log Total Wages	Isolated Entry	0.0602***	(0.0134)	118
Log Total Wages	Clustered Entry	0.0771***	(0.0153)	88
Log Establishments	Isolated Entry	0.0510***	(0.0138)	118
Log Establishments	Clustered Entry	0.0590***	(0.0146)	88
Log Median HH Income	Isolated Entry	0.0200***	(0.0074)	118
Log Median HH Income	Clustered Entry	0.0350***	(0.0077)	88
Unemployment Rate	Isolated Entry	0.1508	(0.1177)	118
Unemployment Rate	Clustered Entry	0.0483	(0.1719)	88
Log Building Permits	Isolated Entry	0.2089***	(0.0583)	118
Log Building Permits	Clustered Entry	0.2508***	(0.0609)	88

* p<0.10, ** p<0.05, *** p<0.01

Isolated Entry: 0 additional DCs within 5 years of first entry.

Clustered Entry: 1+ additional DCs within 5 years of first entry.

Table 8: CS-DID Electricity Heterogeneity: Spillover (Median Split) (Never Treated)

Outcome	Spillover	ATT	SE	N Treated
Residential Price	Low Price Spillover (≤64 Counties)	0.0358*	(0.0189)	168
Residential Price	High Price Spillover (>64 Counties)	-0.0241	(0.0173)	131
Commercial Price	Low Price Spillover (≤64 Counties)	0.0361*	(0.0205)	168
Commercial Price	High Price Spillover (>64 Counties)	-0.0290	(0.0211)	131
Industrial Price	Low Price Spillover (≤64 Counties)	0.0664**	(0.0283)	168
Industrial Price	High Price Spillover (>64 Counties)	-0.0008	(0.0299)	131
Total Price	Low Price Spillover (≤64 Counties)	0.0501**	(0.0225)	168
Total Price	High Price Spillover (>64 Counties)	-0.0539*	(0.0296)	131

* p<0.10, ** p<0.05, *** p<0.01

Spillover = unique other counties sharing any utility with focal county

Figures

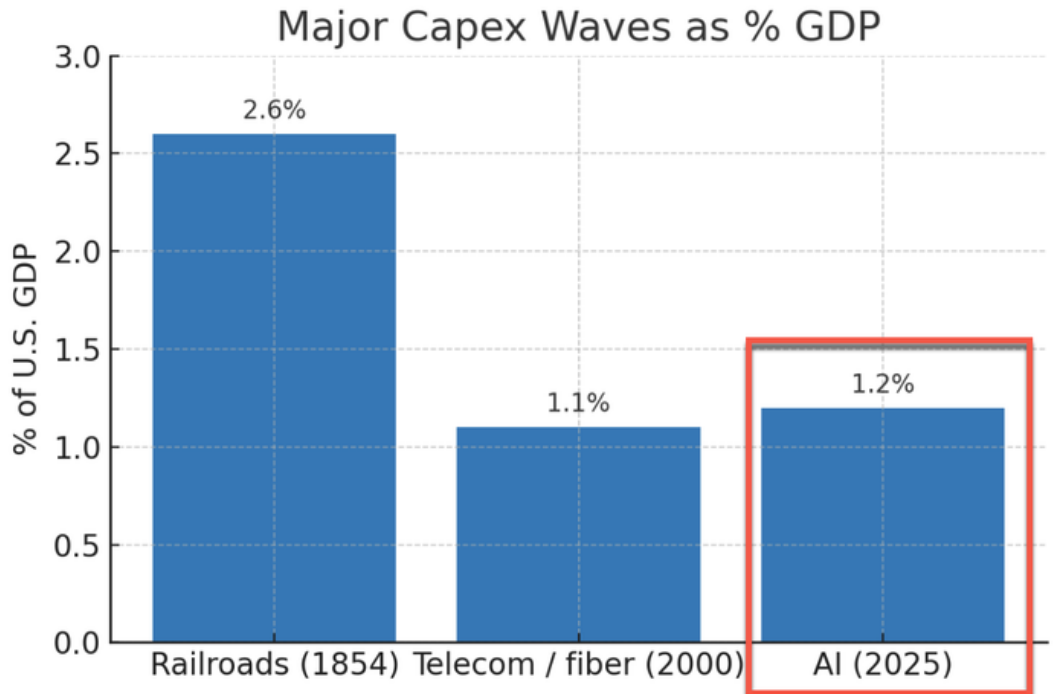


Figure 1: Major Capital Expenditure Waves as Share of U.S. GDP

Notes: This figure compares AI-related capital expenditure to historical infrastructure investment waves including railroads, canals, and fiber optics. Source: Kedrosky and Velu (2024).

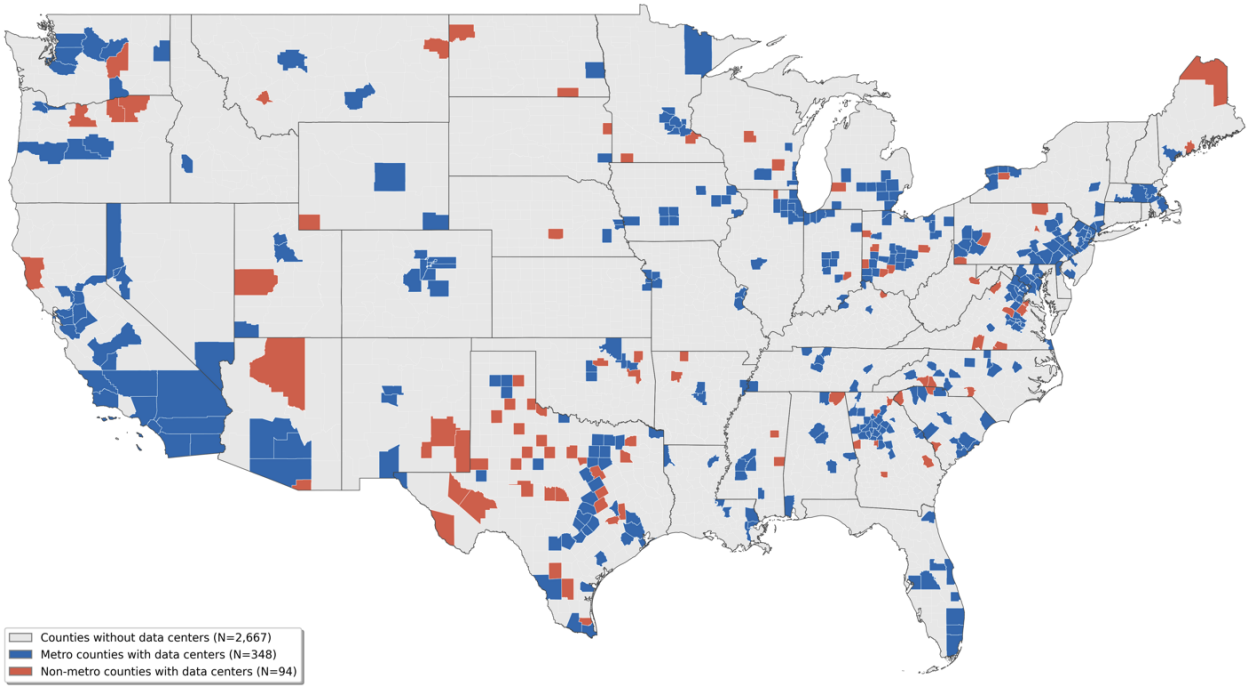


Figure 2: Geographic Distribution of Data Center Counties

Notes: Map shows continental U.S. counties with and without data centers as of 2024Q4. Of 442 counties with data centers, 348 (79%) are metropolitan and 94 (21%) are non-metropolitan. Metro classification based on OMB core-based statistical area delineations.

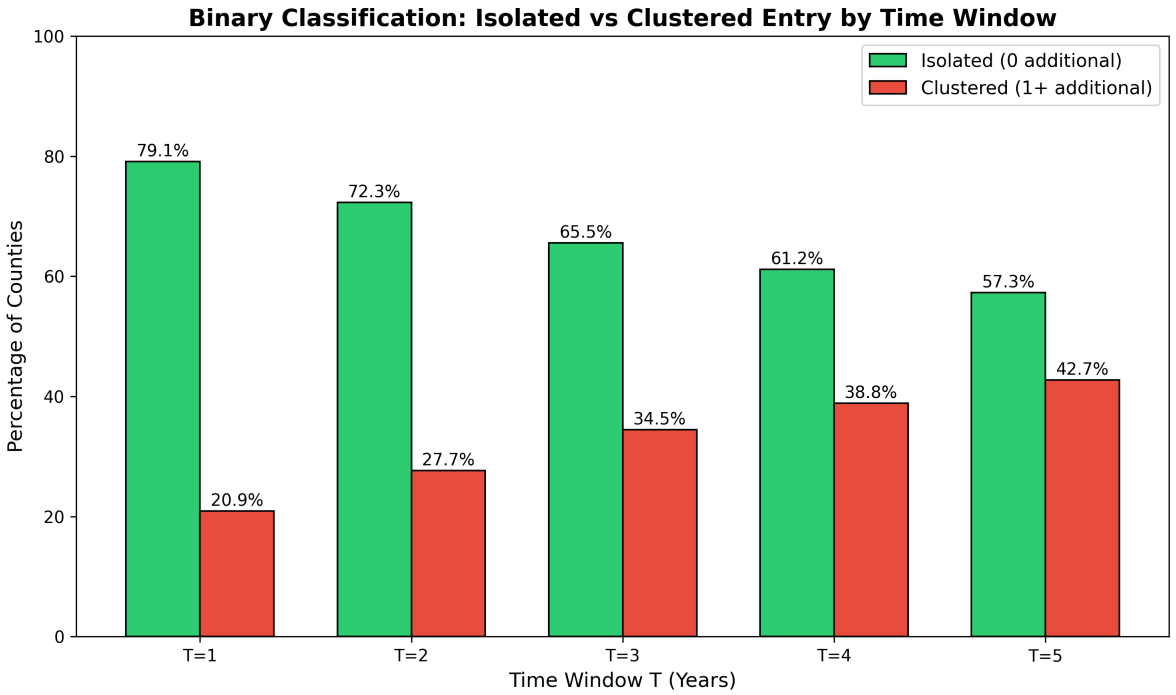


Figure 3: Distribution of Treated Counties by Entry Clustering Classification

Notes: Distribution of treated counties classified as Isolated Entry (0 additional data centers within T years of first entry) versus Clustered Entry (1+ additional data centers within T years). We select T=5 years as our primary specification, yielding 118 isolated counties (57.3%) and 88 clustered counties (42.7%).

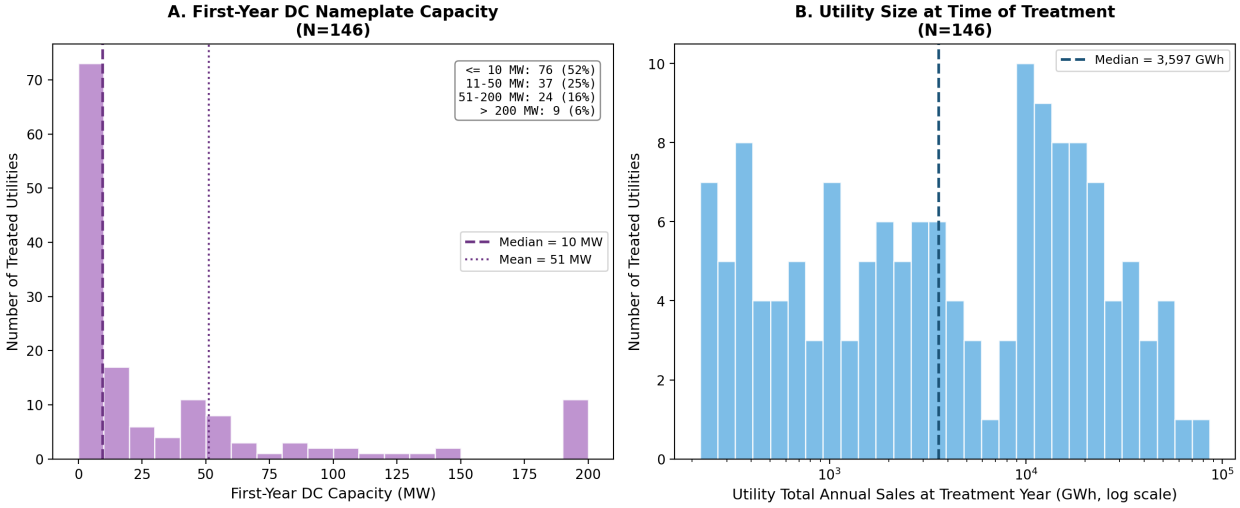
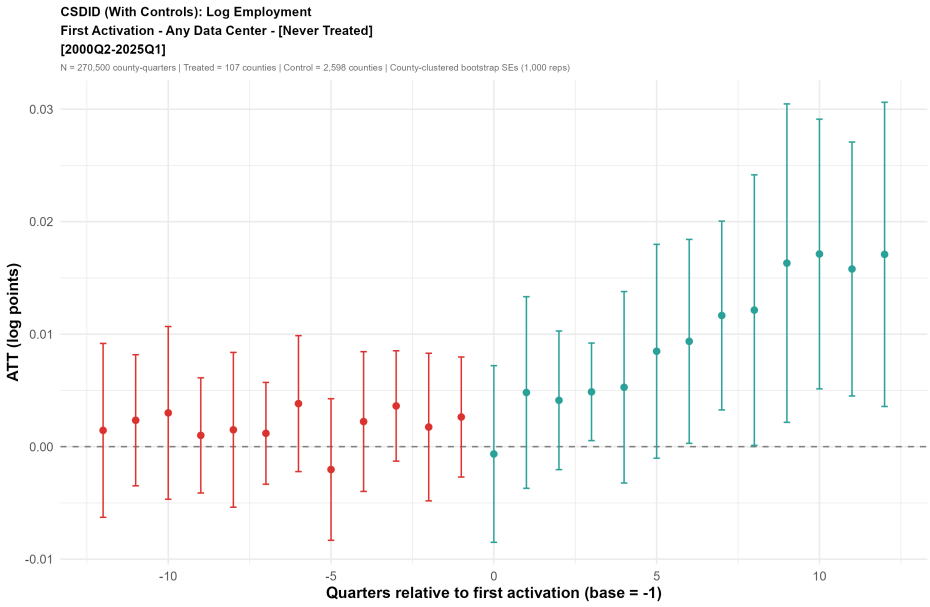
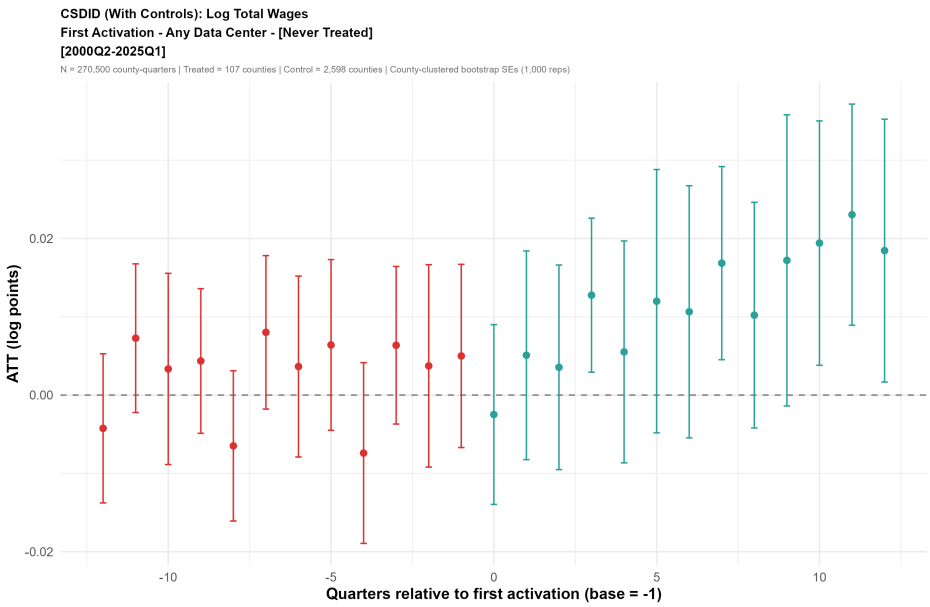


Figure 4: First-Entry Data Center Capacity and Utility Size

Notes: Panel A shows the distribution of first-year data center nameplate capacity (MW) across treated counties. The median first entry is 10 MW, and over 55% of first entries are 10 MW or below. Panel B shows the distribution of total annual electricity sales (GWh, log scale) for the utility serving each treated county at the time of first data center activation. The median utility sells approximately 3,600 GWh per year at the time of first data center entry. At 50% effective utilization of nameplate capacity (accounting for PUE overhead and ramp-up), a 10 MW data center consumes approximately 44 GWh annually—roughly 1.2% of the median utility's total load.



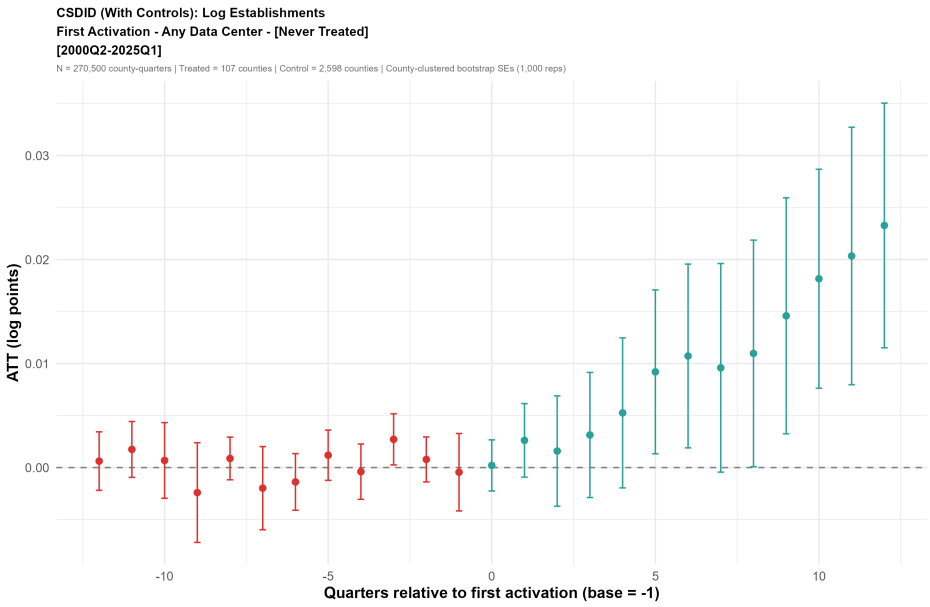
(a) Log Employment



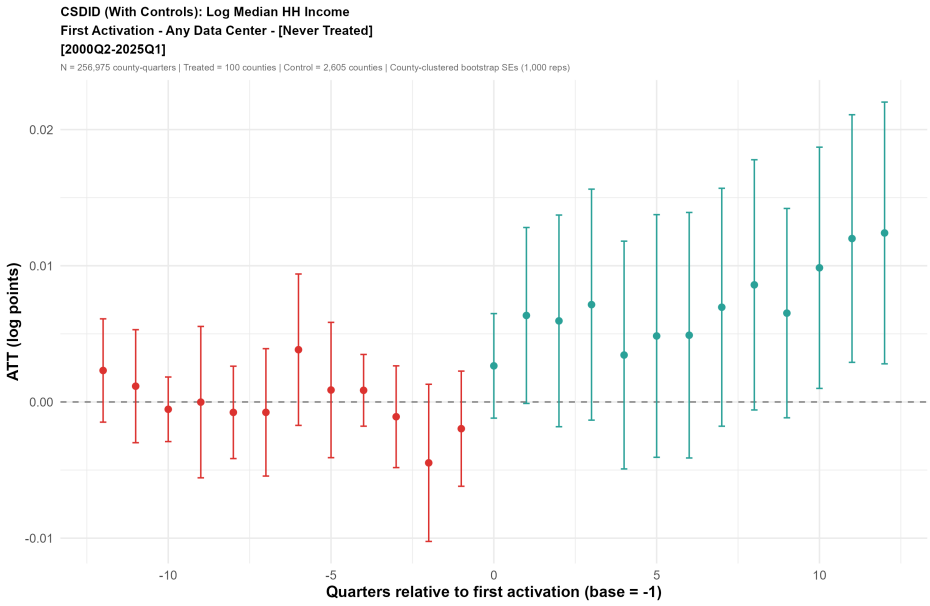
(b) Log Total Wages

Figure 5: Baseline Event Study Estimates (Part 1 of 3)

Notes: Event study estimates from Callaway-Sant'Anna DiD. Control group: never-treated counties. Shaded areas: 95% CI. Event time 0 = quarter of first data center activation.

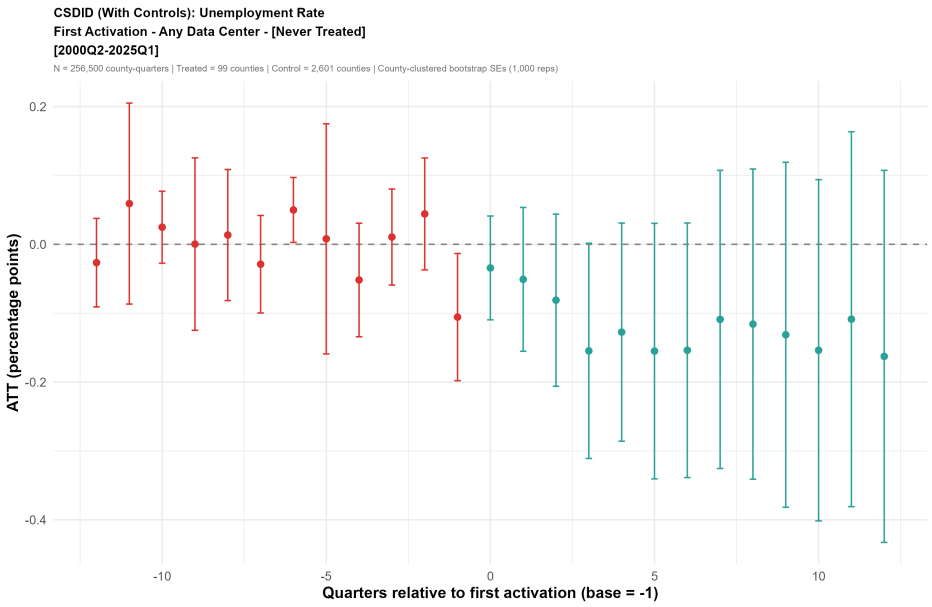


(a) Log Establishments

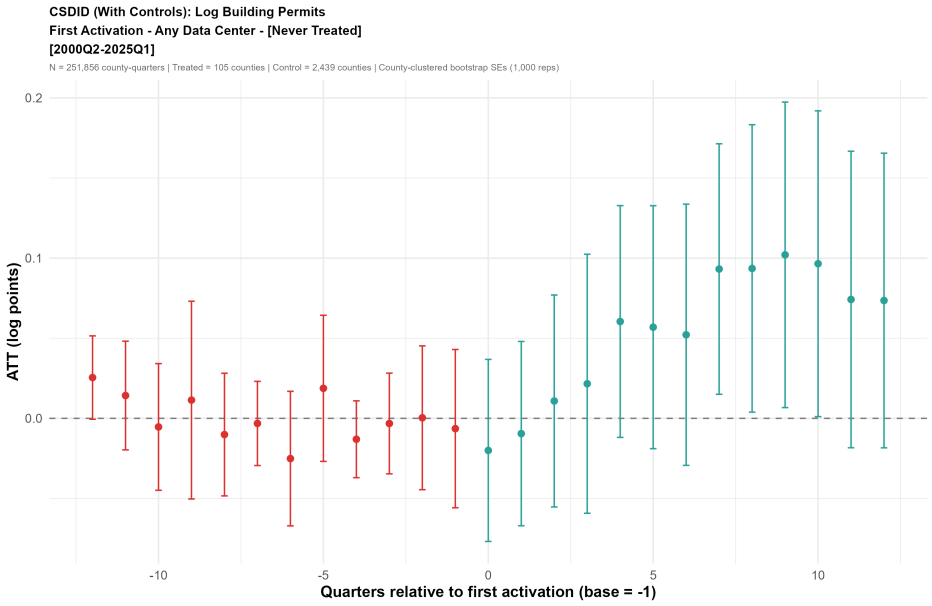


(b) Log Median HH Income

Figure 6: Baseline Event Study Estimates (Part 2 of 3)

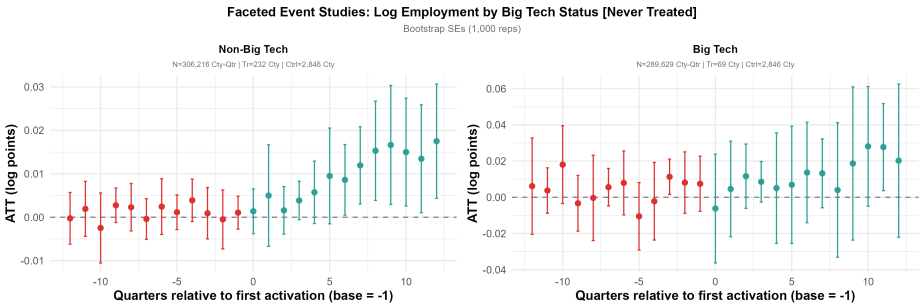


(a) Unemployment Rate

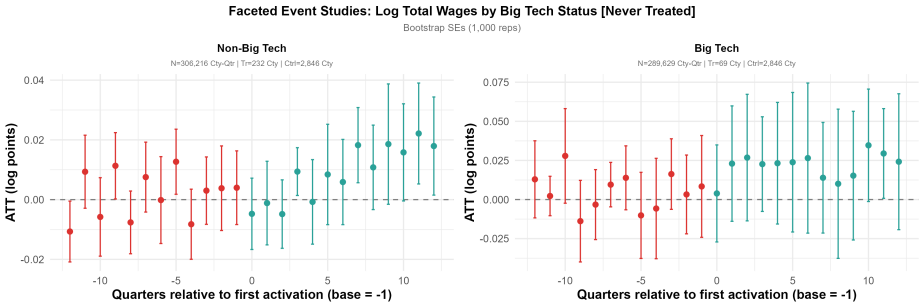


(b) Log Building Permits

Figure 7: Baseline Event Study Estimates (Part 3 of 3)



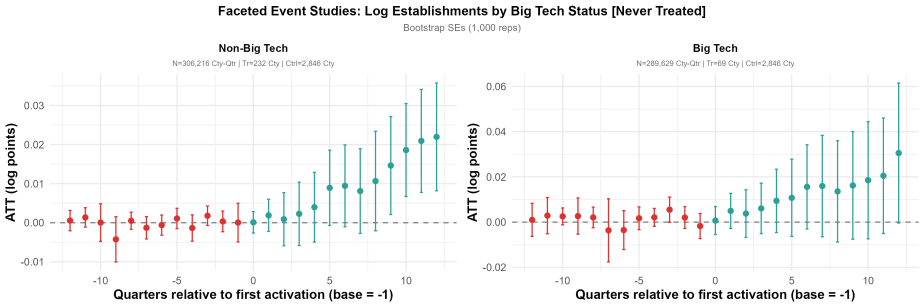
(a) Log Employment



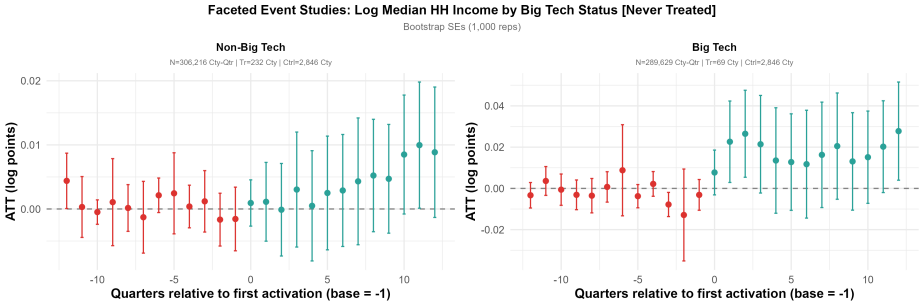
(b) Log Total Wages

Figure 8: Event Study by Operator Type (Part 1 of 3)

Notes: Big Tech includes Amazon, Google, Microsoft, Meta, Apple, and Oracle. Control: never-treated.

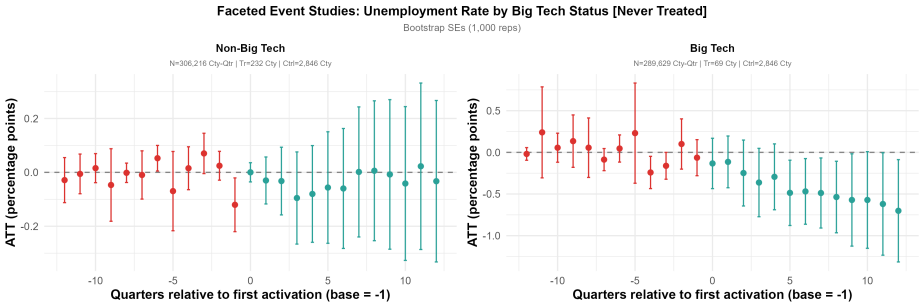


(a) Log Establishments

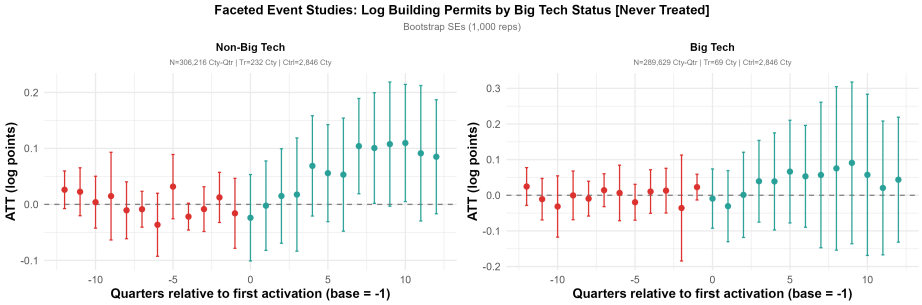


(b) Log Median HH Income

Figure 9: Event Study by Operator Type (Part 2 of 3)

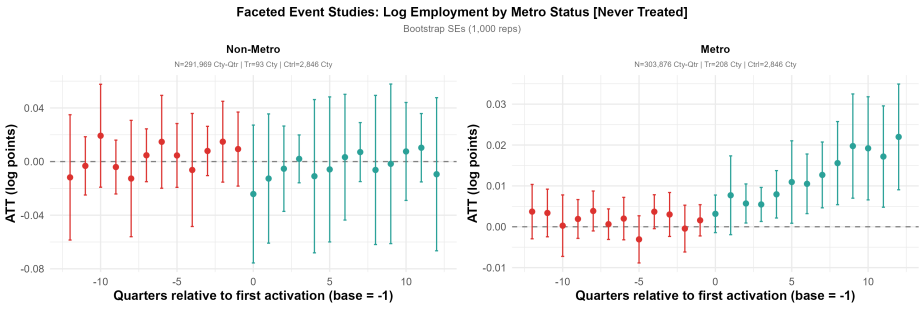


(a) Unemployment Rate

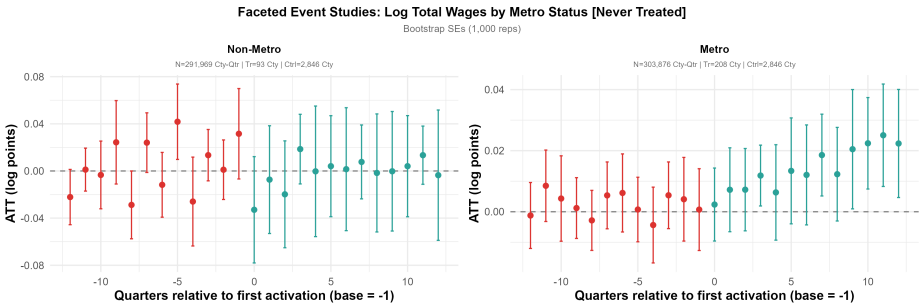


(b) Log Building Permits

Figure 10: Event Study by Operator Type (Part 3 of 3)



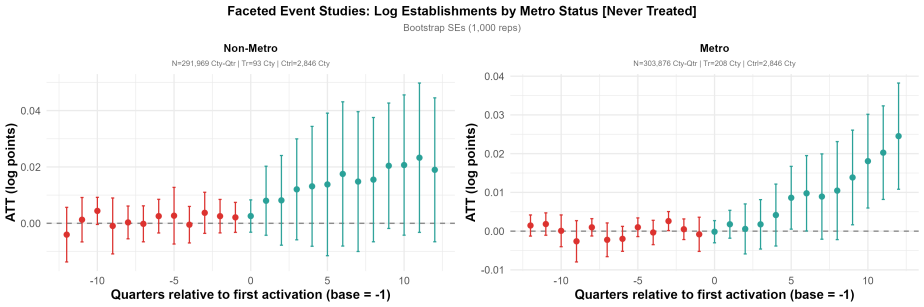
(a) Log Employment



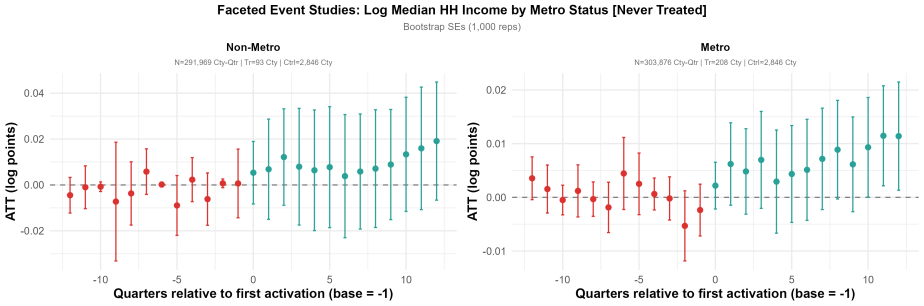
(b) Log Total Wages

Figure 11: Event Study by Metro Status (Part 1 of 3)

Notes: Metro classification based on OMB CBSA delineations. Control: never-treated.



(a) Log Establishments

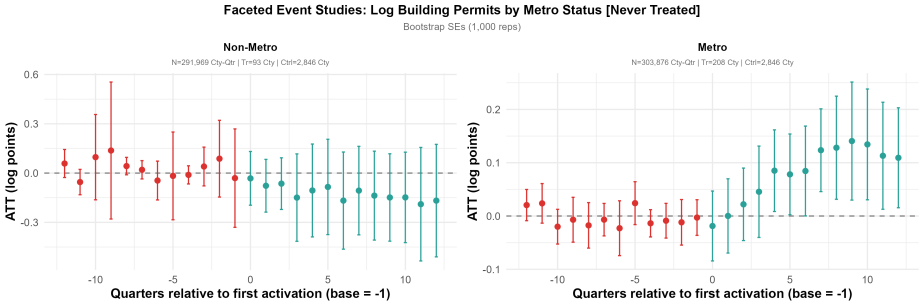


(b) Log Median HH Income

Figure 12: Event Study by Metro Status (Part 2 of 3)

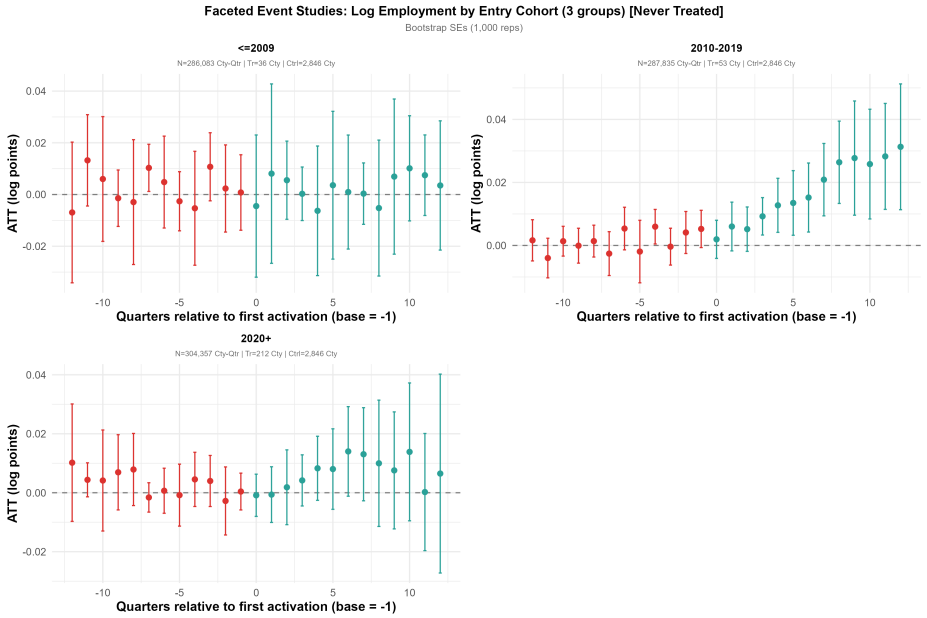


(a) Unemployment Rate

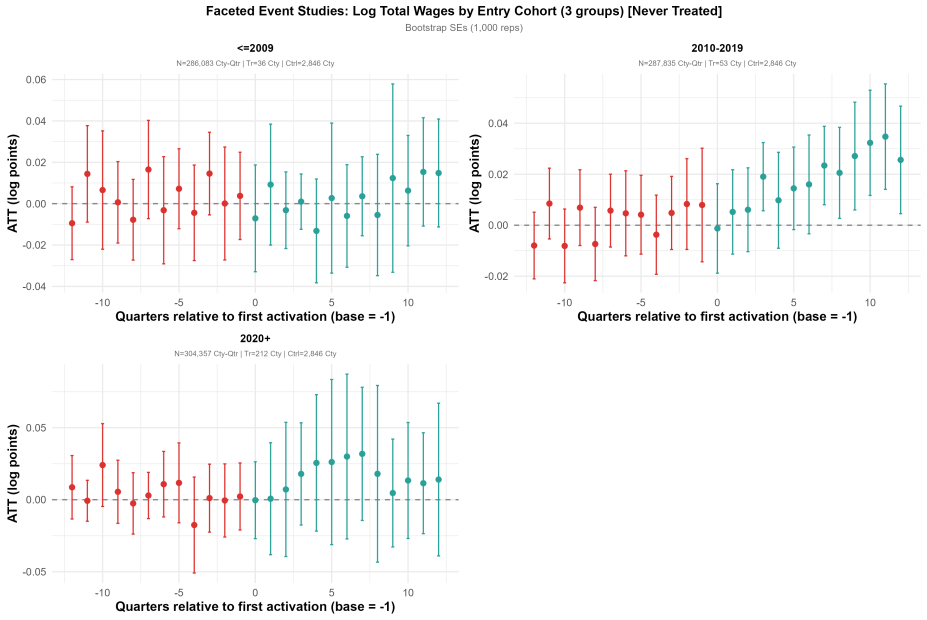


(b) Log Building Permits

Figure 13: Event Study by Metro Status (Part 3 of 3)



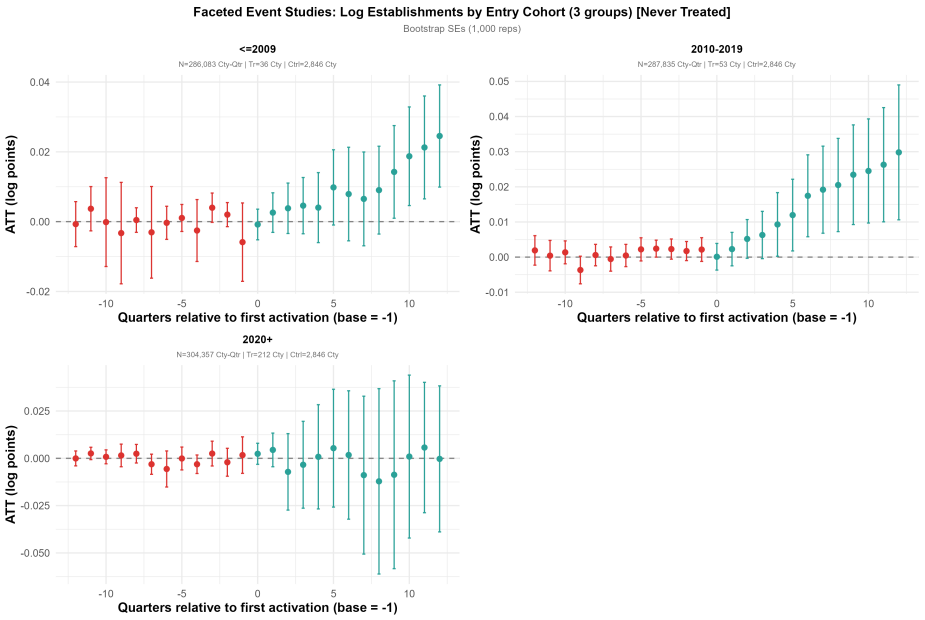
(a) Log Employment



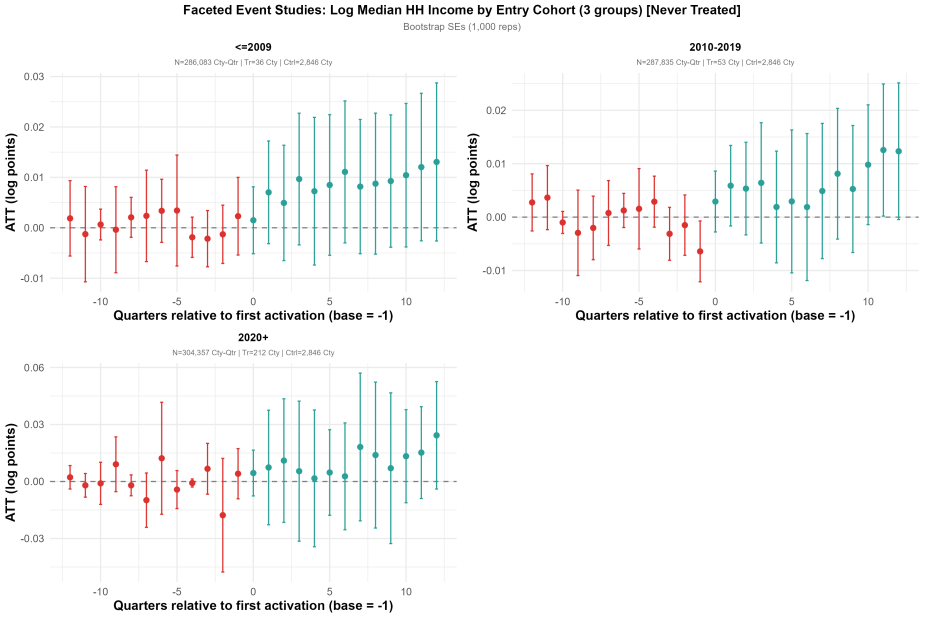
(b) Log Total Wages

Figure 14: Event Study by Entry Cohort (Part 1 of 3)

Notes: Cohorts: Pre-2010, 2010–2019, Post-2020. Control: never-treated.

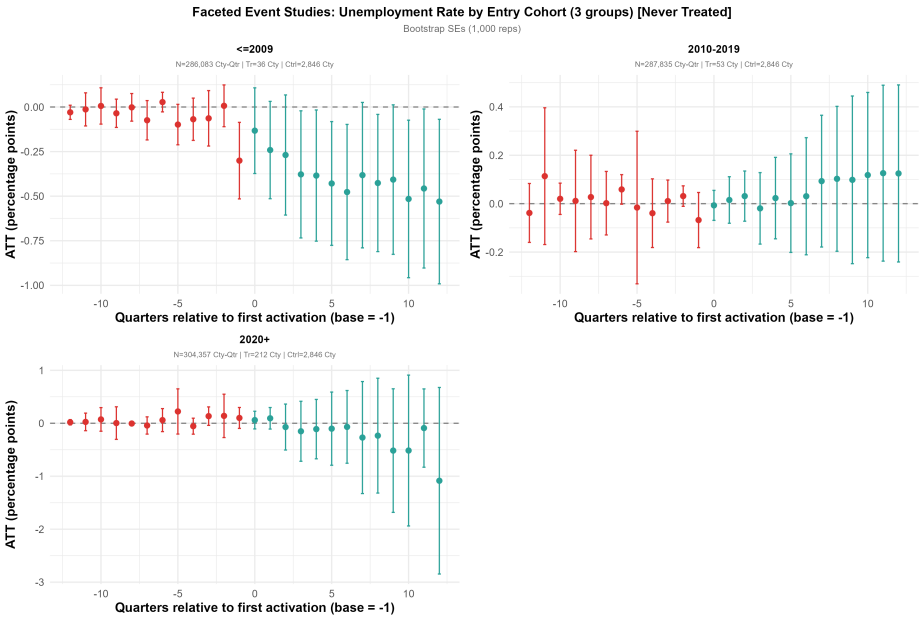


(a) Log Establishments

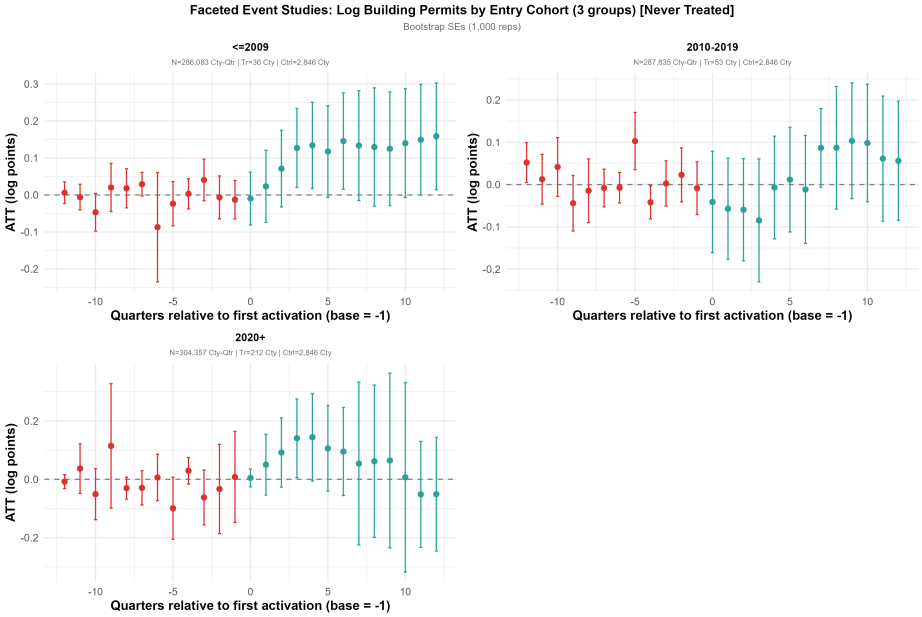


(b) Log Median HH Income

Figure 15: Event Study by Entry Cohort (Part 2 of 3)

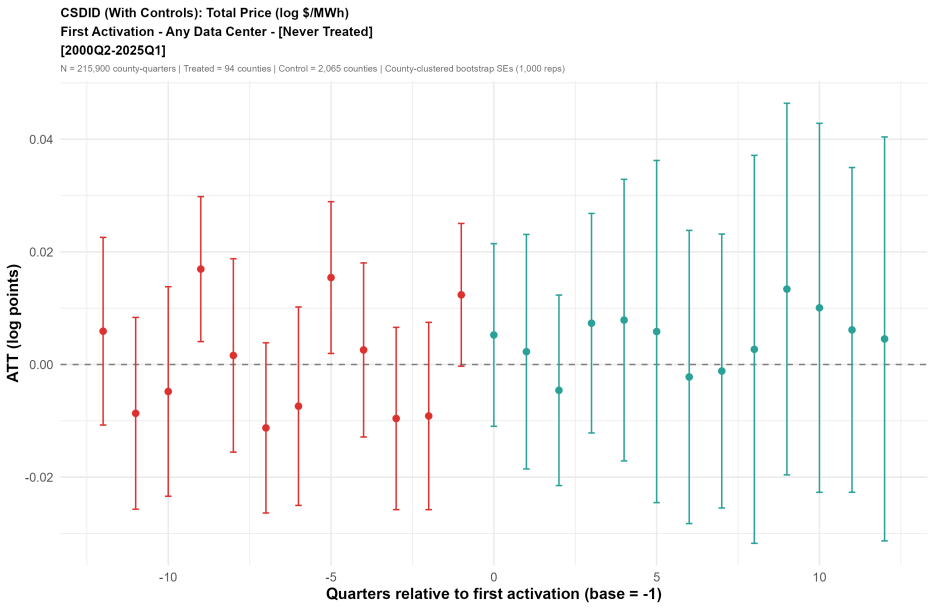


(a) Unemployment Rate

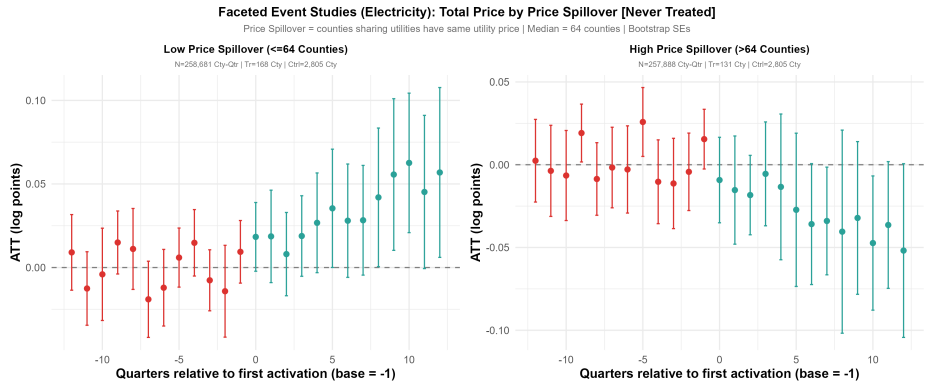


(b) Log Building Permits

Figure 16: Event Study by Entry Cohort (Part 3 of 3)



(a) Baseline: Log Total Electricity Price



(b) By Spillover Exposure: Log Total Electricity Price

Figure 17: Electricity Price Event Studies

Notes: (a) Baseline event study for log total electricity price. (b) Faceted by spillover exposure; low spillover ($\leq$64 shared counties) provides cleaner identification. Control: never-treated. Prices from EIA-861M.

Appendix A: Adjacent County Analysis

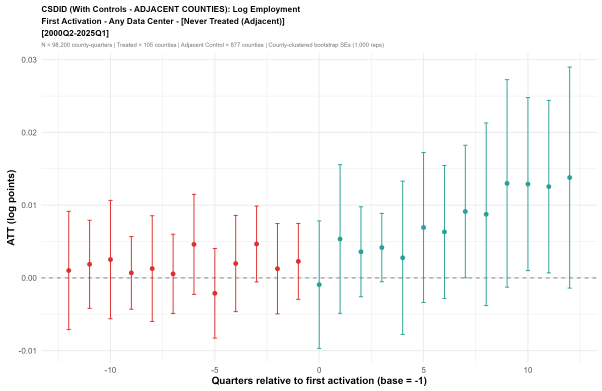
As a robustness check, we restrict the control group to counties that are geographically adjacent to treated counties. This uses the U.S. Census Bureau's County Adjacency File to identify counties sharing a physical border with any treated county. Approximately 102 treated counties and 880 adjacent control counties remain after applying sample restrictions. This design provides a more localized comparison, though effects may be understated if spillovers extend to neighboring counties.

Table A1: CS-DID Baseline Results - Adjacent Controls (Never Treated (Adjacent))

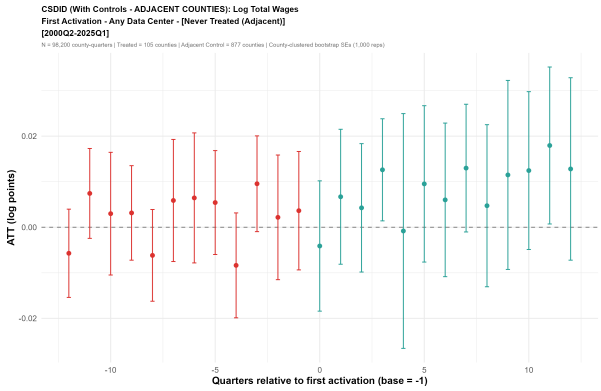
Outcome	ATT	SE	95% CI Lower	95% CI Upper
Log Employment	0.0317***	(0.0120)	0.0083	0.0552
Log Total Wages	0.0436***	(0.0131)	0.0179	0.0692
Log Establishments	0.0472***	(0.0122)	0.0233	0.0711
Log Median HH Income	0.0184***	(0.0053)	0.0081	0.0288
Unemployment Rate	-0.1064	(0.1518)	-0.4038	0.1910
Log Building Permits	0.1546***	(0.0498)	0.0570	0.2521

* p<0.10, ** p<0.05, *** p<0.01

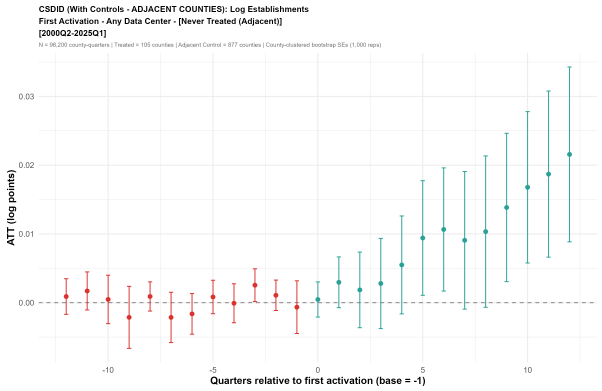
Sample: 2000Q2-Present, 280 treated, 803 adjacent controls



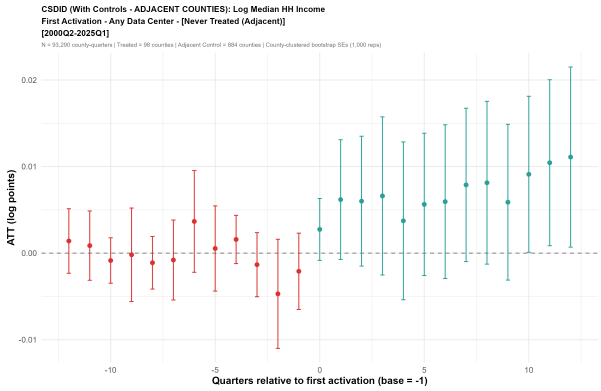
(a) Log Employment



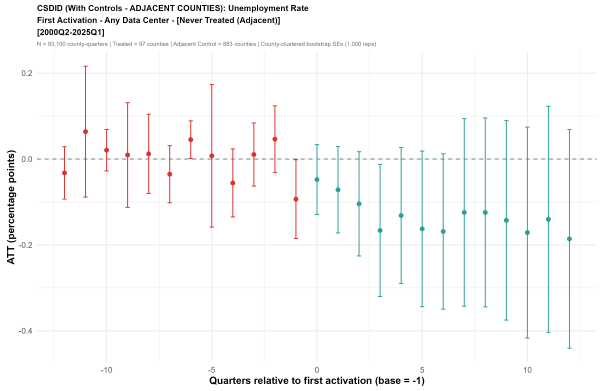
(b) Log Total Wages



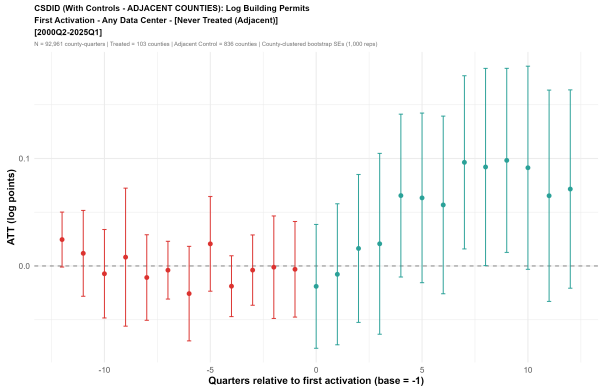
(c) Log Establishments



(d) Log Median HH Income



(e) Unemployment Rate



(f) Log Building Permits

Figure A1: Adjacent County Event Studies

Notes: Control group: never-treated adjacent counties only.

Appendix B: Not-Yet-Treated Control Group

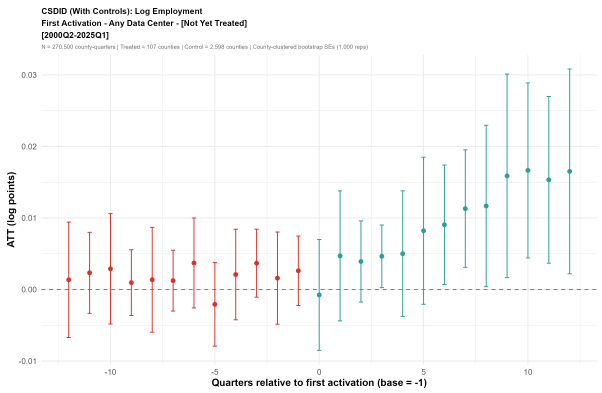
Our main results use never-treated counties as controls. Here we present results using not-yet-treated controls as a robustness check. Under this specification, the control group for counties treated in quarter g includes all counties not yet treated by each period t—both never-treated and future-treated counties. This expands the control pool but requires the assumption that future-treated counties were not already affected by anticipation of treatment.

Table B1: CS-DID Baseline Results (Not Yet Treated)

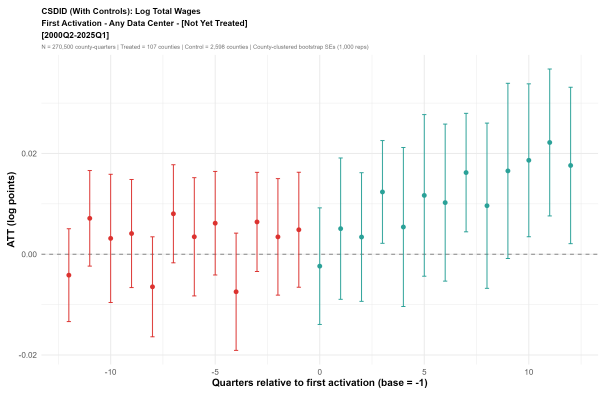
Outcome	ATT	SE	95% CI Lower	95% CI Upper
Log Employment	0.0340***	(0.0106)	0.0133	0.0548
Log Total Wages	0.0487***	(0.0119)	0.0254	0.0720
Log Establishments	0.0461***	(0.0113)	0.0240	0.0681
Log Median HH Income	0.0186***	(0.0047)	0.0094	0.0278
Unemployment Rate	-0.0347	(0.1539)	-0.3364	0.2669
Log Building Permits	0.1571***	(0.0508)	0.0575	0.2566

* p<0.10, ** p<0.05, *** p<0.01

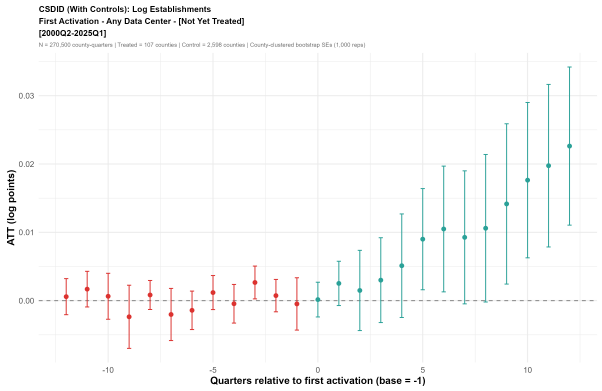
Sample: 2000Q2-Present, 301 treated counties



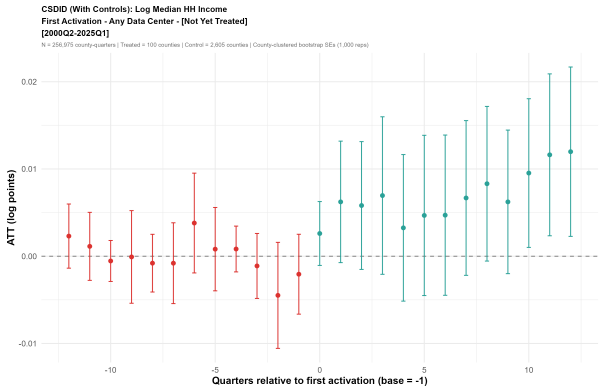
(a) Log Employment



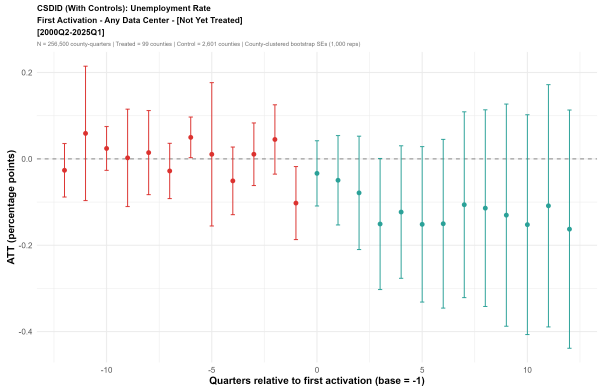
(b) Log Total Wages



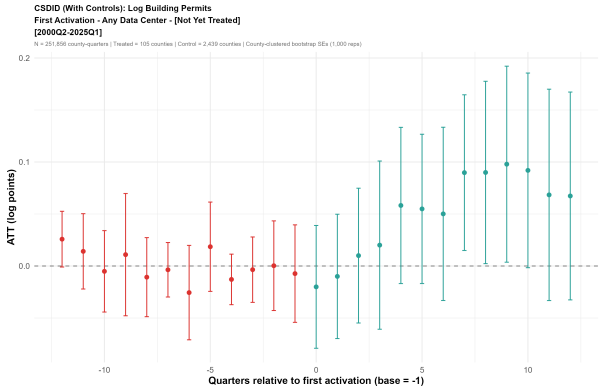
(c) Log Establishments



(d) Log Median HH Income



(e) Unemployment Rate



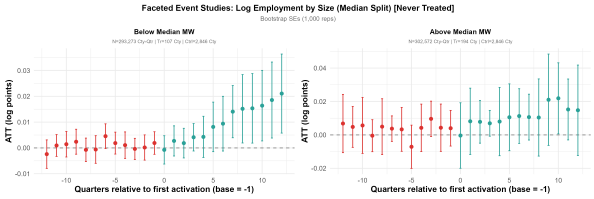
(f) Log Building Permits

Figure B1: Not-Yet-Treated Event Studies

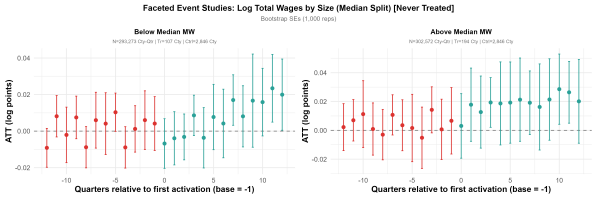
Notes: Control group: not-yet-treated counties.

Appendix C: Additional Heterogeneity Analysis

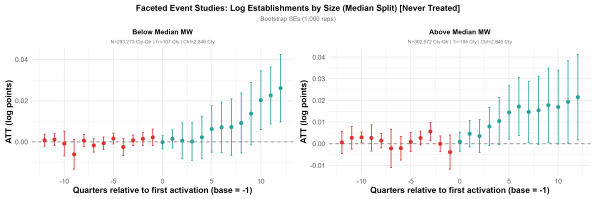
This appendix presents additional heterogeneity analyses by data center size (MW capacity, median split), county population (median split), and RUCC category (Large Metro = RUCC 1; Medium Metro = RUCC 2; Smaller Areas = RUCC 3–9).



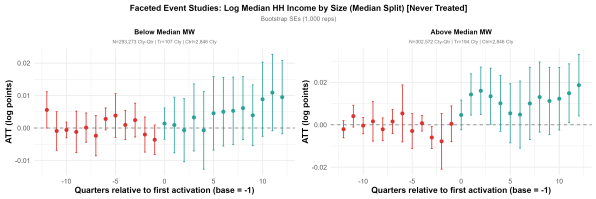
(a) Log Employment



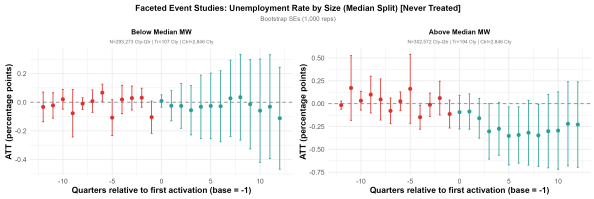
(b) Log Total Wages



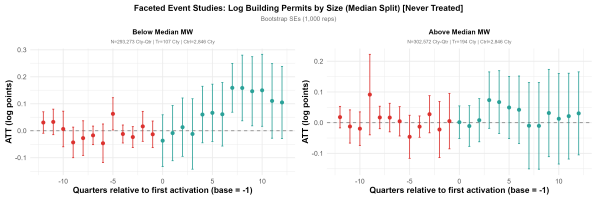
(c) Log Establishments



(d) Log Median HH Income



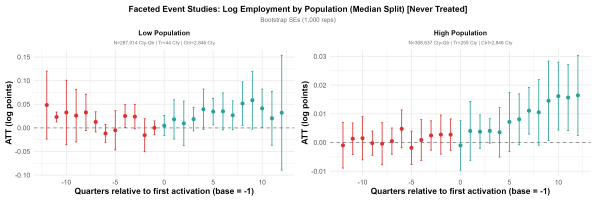
(e) Unemployment Rate



(f) Log Building Permits

Figure C1: Heterogeneity by Data Center Size (MW)

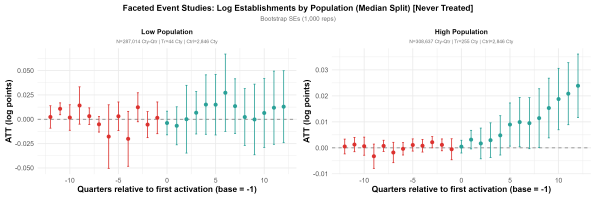
Notes: Median split among treated counties. Control: never-treated.



(a) Log Employment



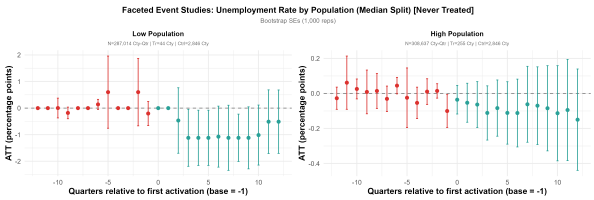
(b) Log Total Wages



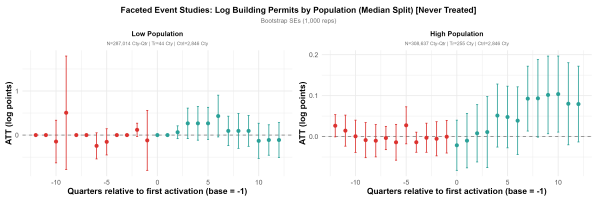
(c) Log Establishments



(d) Log Median HH Income



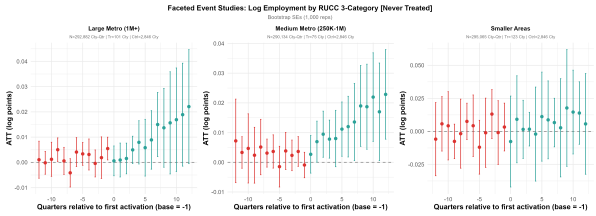
(e) Unemployment Rate



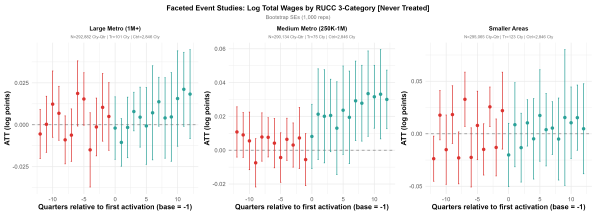
(f) Log Building Permits

Figure C2: Heterogeneity by County Population

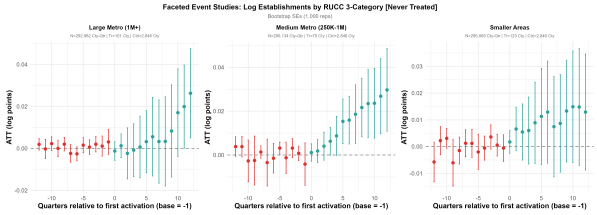
Notes: Median split among treated counties. Control: never-treated.



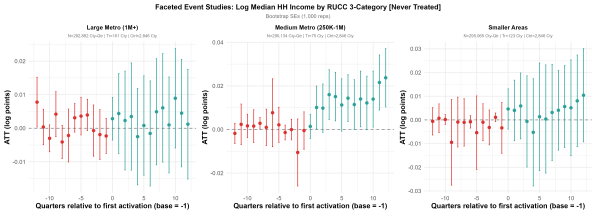
(a) Log Employment



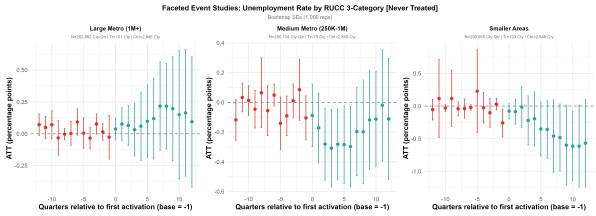
(b) Log Total Wages



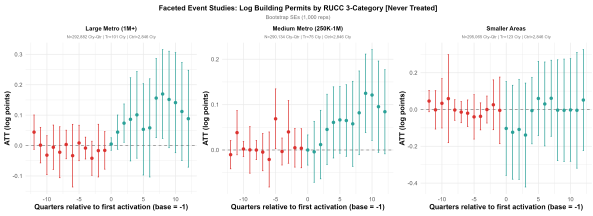
(c) Log Establishments



(d) Log Median HH Income



(e) Unemployment Rate



(f) Log Building Permits

Figure C3: Heterogeneity by RUCC Category

Notes: Large Metro = RUCC 1; Medium Metro = RUCC 2; Smaller = RUCC 3–9.