Subsidizing the Cloud: U.S. State Incentives to Data Centers

Antonio Gargano[*]
University of Houston
and University of Melbourne

Marco Giacoletti[†]
University of Notre Dame

First Draft: November 2025
This Draft: February 2026

Abstract

Over the past two decades, many U.S. states have introduced tax incentives to attract data center investment. We provide causal evidence that these incentives impact state-level data center construction. However, effects are concentrated among large, high-power facilities used for cloud computing and, more recently, large language models. Smaller data centers remain largely unaffected. Using a quantitative model, we show that incentives significantly increase project net present value and shift the geographic distribution of hyperscale development across states. However, we find no clear empirical evidence that data centers stimulate local tech-related employment growth.

Keywords: Infrastructure Investments, Data Centers, Tax Incentives, Job Creation
JEL Classification: G38, H25, R3

[*]C.T. Bauer College of Business. Email: agargano@bauer.uh.edu. Honorary Research Fellow, University of Melbourne: antonio.gargano@unimelb.edu.au

[†]University of Notre Dame, Mendoza College of Business. Email: mgiacole@nd.edu

1 Introduction

Local governments use tax incentives to stimulate infrastructure development, expecting benefits to local economic growth and job markets.[1] However, existing research shows that incentive programs may fail to attract productive investments, and that new infrastructure may have limited impact on local economies.[2] This is particularly a concern for modern infrastructure, which frequently produces goods and services that are not consumed locally, is operated with minimal labor, and responds to forces and incentives that are beyond the control of local regulators. A case in point are data centers, hosting large numbers of servers, which are the backbone of the internet, high-speed server networks, cloud computing, and most recently Large Language Model (LLM) workloads. Over the past two decades, 38 U.S. states have introduced tax incentives for data center investments. To our knowledge, it is still not well established whether these incentives are shaping the development of data centers across states, and, if so, what the implications are for the data center industry and for local employment growth.

In this paper, we examine whether U.S. state-level tax incentives have been successful in attracting data center development, assess the economic costs of the incentives and the value delivered to data center investors, and finally explore the implications for local job markets.

We first estimate the causal effect of the incentive policies on development using a staggered difference-in-differences design and methods that address issues related to heterogeneous treatment effects across units over time.[3] We find that state policies have, on average, more than doubled the scale of data center development, but only for large data centers with high electric power demand, especially hyperscale data centers designed to support high-speed cloud computing and modern LLM and AI technology. We show that this result is driven by the structure of the incentives, which favor large-scale projects.

[1] For example, Amazon's search for a second headquarters location in 2018 involved over 200 cities offering tax breaks, infrastructure improvements, and other incentives. Nevada provided $1.3 billion in tax breaks and incentives over 20 years for the construction of Tesla Gigafactory.

[2] See the literature review at the end of this section.

[3] We follow the approaches in Cengiz et al. (2019), Sun and Abraham (2021), and Callaway and Sant'Anna (2021).

We find that there is substantial heterogeneity across states in the value delivered by incentives to data center investors and operators. However, in several states that are major data center markets, these incentives generate millions of dollars in annual operating expense savings for individual facilities.

Finally, we consider the implications for the labor market and focus on the impact of data centers on the creation of long-term occupations in the tech sector, including information technology, professional services, and tech-related manufacturing. We find that estimates of these effects are noisy and statistically insignificant, or quantitatively small. These results do not imply that there are no local economic benefits from data centers. The construction of data centers likely boosts temporary local employment in the construction sector and leads to a concurrent short-term increase in local economic growth. However, tax incentives do not take the form of short-term subsidies to development, but are instead designed as long-term measures, reducing asset operating costs through sales tax rebates. Thus, we believe that long-term effects, aligned with the duration of states' fiscal commitments, must be taken into account. Moreover, regulators have argued that attracting large data center projects is an instrument for accelerating growth in their state's tech sector.[4]

Our analysis is based on new high-quality data from S&P Global, which reports the precise geolocation of U.S. data centers, their year of construction, and detailed information on size, power requirements, operations, and ownership. The dataset also provides information on data centers that, as of Q2 2025, are under construction or planned.

We focus our analysis on facilities completed between 1994 and 2025, and on the three main categories of data centers available in the dataset: retail, wholesale, and hyperscale. It is important to note that in this sector, asset size is defined by the electric power capacity that a data center can reliably allocate to its servers. This capacity, known as Uninterruptible Power Supply (UPS), determines the amount of computing power the facility can support.

Retail data centers focus on network services, and are owned by either real estate investors

[4]See public statements by Governors of Michigan, Mississippi, and Indiana.

(who rent out electric capacity to the servers of tenants) or IT companies. They have the smallest scale, with UPS capacity of 4 megawatt on average.[5] Wholesale data centers are used for cloud computing, and have a broad range of sizes. Smaller facilities offer services that complement those provided by retail data centers, whereas larger ones offer services that support or complement the operations of the largest, hyperscale data centers. Wholesale data centers in our dataset are frequently owned by real estate investors or LLCs. Finally, hyperscale data centers are, as mentioned, the largest assets, with a mean size of 33 UPS megawatt. The earliest facilities belonging to this category were completed in the late 2000s and early 2010s. Hyperscale data centers have then expanded rapidly in the 2020s, driven by the considerable computing needs associated with LLMs and other AI applications. They are typically owned and operated by major technology firms.

We combine information on data center development with manually collected information on tax incentive laws (e.g., the year of enactment, eligibility requirements, and scope of benefits) from state-maintained online databases of codified statutes. While the earliest of these provisions was implemented in 2000, most legislations were introduced in the 2010s. As mentioned above, nearly all state laws provide sales and use tax incentives for newly built data centers that meet minimum thresholds for size or capital investment. The focus on sales tax incentives is likely motivated by the cost structure of data centers, which incur high operating expenses, primarily for electricity, and high capital expenditures for replacing electrical equipment and infrastructure. Since these inputs are typically subject to state sales and use taxes, exemptions can meaningfully improve operating margins and generate value for investors and operators.

In our staggered difference-in-differences design, we find that the implementation of tax incentives in a state increases the likelihood that at least one new hyperscale data center is completed in a given year by 26%. The number of new hyperscale data centers completed each year more than doubles, and the total UPS power for new completed hyperscale data

[5]For a reference, a typical domestic electricity user in the U.S. has access to 24 kW. 4 MW is 167 times more. In addition, a domestic user is unlikely to use available power continuously, while the UPS used by data centers is likely to be continuous.

centers increases by more than 250%. We find significant effects for wholesale data centers, but magnitudes are approximately one-third of those estimated for hyperscale. Treatment effects are small and not statistically significant for retail data centers. Our regressions include state and year fixed effects, and time-varying state controls measuring electricity costs and demand at the state level. The inclusion of these controls has virtually no impact on our treatment effects estimates. The results are robust across various specifications, including the approaches developed by Cengiz et al. (2019), Sun and Abraham (2021), and Callaway and Sant'Anna (2021).

Thus, tax incentives substantially impact data center development within states. This outcome is not something to take for granted, as evidence on the effectiveness of tax incentive programs is often mixed.[6] Additional tests show that incentives lead to the development of larger, higher-UPS facilities geared toward cloud computing and serving a single major operator. These new facilities are also slightly less likely to include on-site offices and slightly more likely to be located in low-population density areas. These shifts are driven by the increase in the share of hyperscale facilities.

Details on the provisions of the laws allow us to isolate the mechanism behind the heterogeneous treatment effects across facility sizes. Although differences in technology and operating costs may matter, the primary mechanism is related to investment scale. The statutes effectively exclude smaller assets. Back-of-the-envelope calculations show that, despite pre-dating the hyperscale boom, most laws set minimum requirements that would have qualified the vast majority of hyperscale data centers in 2020, but fewer than 20% of retail data centers.

We then turn to quantifying the operating cost savings generated for data center operators by the laws. Our analysis focuses on sales and use tax incentives, for which we collect detailed information on the coverage and exemptions embedded in each statute. We merge these statutory details with facility-level data on technology and operations from our main dataset,

[6] See, for example, the evidence from studies examining the effects of Opportunity Zone tax incentives, which were created by the 2017 Tax Cuts and Jobs Act (Chen, Glaeser, and Wessel, 2023, Arefeva et al., 2024, and Corinth and Feldman, 2024).

together with supplemental industry information. We then estimate average tax savings per megawatt of UPS capacity in each state. Although the estimates are indicative and depend on assumptions regarding electricity usage and power efficiency, they provide a comparable benchmark across states

We find that several states with large data center markets provide substantial incentives. For example, the laws in Texas, Iowa, and South Carolina generate savings in operating costs of between $80,000 and $90,000 per year and per megawatt of UPS capacity, while savings in North Carolina are close to $70,000. Some emerging markets, such as Indiana and Minnesota, offer savings that approach or exceed $100,000 per megawatt.

We extend this analysis by developing a simple quantitative model that formalizes the discrete choice problem of investing in data center development across states, incorporating the option to delay investment. In the model, operating income dynamics are driven by state variables capturing local electricity costs as well as electrical equipment and infrastructure costs. The model also accounts for the effects of state-level tax incentives on operating income. We calibrate the model so that the distribution of investment probabilities across states matches the distribution of new hyperscale projects under construction or planned as of 2025.

We then conduct two counterfactual exercises. In the first, we sequentially remove incentives for a single state and recompute the spatial distribution of hyperscale development. We find that unilaterally removing incentives would lead, for nearly all states, to a drop in investment probabilities exceeding 50%, which we interpret as a corresponding contraction in their expected hyperscale pipeline. These magnitudes are broadly consistent with our empirical findings and help illustrate the potential effects of repealing a state's incentive program, as has been proposed in Georgia and, most recently, Illinois.

In the second exercise, we simultaneously remove incentives from all states. We find that this reduces the net present value of hyperscale investments by approximately $2 million per megawatt in states offering substantial tax savings. We also observe a significant redistribution of investment activity, with more than 10% of the development pipeline being reallocated across

states and some of the states with the most generous subsidies experiencing drops in investment probabilities exceeding 50%.

Because job creation is frequently cited by state and local governments as a goal of data center incentive policies, we next examine the effects of data centers on local labor markets, focusing on the creation of new tech-related jobs. We merge the data center records with employment histories from RevelioLab, which compiles individual job profiles from online sources and links them to specific employers and locations. From these data, we construct several alternative measures of the number of new tech jobs initiated each year in each census place (city, town, or unincorporated community), and determine the exposure of each census place to data center development. Our definitions of tech jobs are intentionally broad, spanning information technology, professional and technical services, and manufacturing.

Our analysis focuses on Arizona, Iowa, Minnesota, North Carolina, Nevada, Ohio, South Carolina, Texas, and Virginia. We select these states because they both offer some of the most valuable incentive packages for new data center development (Figure 5) and have a significant concentration of hyperscale and wholesale facilities as of 2025. Since our empirical strategy relies on past data center development, these states provide a sufficient number of observations to support meaningful inference (Figure A.3).

Within each state, we compare tech job growth across census places that are, and are not, exposed to new data center development within a 10-mile radius. This radius is intended to capture the broader labor market spillovers generated by new facilities. Tech jobs benefiting from proximity to the data center may not just be located directly on site, but also in the surrounding cities and towns.

Our regressions include census place and year fixed effects and estimate the impact of new data center development on the number of individuals starting a new tech job in subsequent years. This specification absorbs place-specific trends in job growth as well as state-level shifts in the broader tech industry, allowing us to isolate the relationship between local development and new tech job formation.

However, the concern remains that data center development is not randomly allocated in space. City features that attract data center development may also lead to the acceleration of local tech job creation over time, or vice versa. To attenuate concerns about confounding factors impacting our results, we use a propensity score matching approach. Our estimation sample is restricted to census places that experienced development within 10 miles, and census places that did not experience any data center development within 10 miles as of 2025, but that are matched with the exposed places based on 2012 log median household income, log median home value, median age of household heads, log total population, poverty rate, and share of population with bachelor or higher degree.

In most states, there are no statistically significant differences in observable demographic characteristics between places exposed to data center development and matched not-exposed places. At the same time, the approach reveals that exposed locations tend to have higher incomes, higher home values, and a larger share of residents with a bachelor's degree. These attributes, and potentially other unobservable factors correlated with them, may drive tech job creation, independent of data center activity. Consequently, any remaining imbalances likely bias our estimates toward finding a positive relationship between data center development and tech job starts.

Across specifications, the estimated effects of data center development on new tech-related jobs are small and statistically insignificant for almost all states. We find some evidence of job creation only in Ohio. A back-of-the-envelope calculation suggests that, in this state, the tax incentive allocated for each new job created is larger than $140,000, and potentially as large as $500,000. Notably, this is likely the most cost-effective program among the states we examine. For the other states, the implied cost per new job easily exceeds $500,000, and, as noted above, the corresponding point estimates of job creation are frequently insignificant.

This paper contributes to the literature that studies investments in infrastructure projects and their real effects. Leduc and Wilson (2012) provide a comprehensive discussion of the trade-offs and challenges faced by public infrastructure programs. Gupta, Van Nieuwerburgh,

and Kontokosta (2022) study private value creation from public infrastructure projects, while Agarwal et al. (2023), Agarwal et al. (2025), and Boudot-Reddy and Butler (2024) explore a broad range of social and financial outcomes impacted by the creation of new infrastructure.

Other recent papers have studied the role of institutional investors in infrastructure (Andonov, Kraussl, and Rauh, 2021, Howell et al., 2023, and Agarwal, Mukhrjee, and Naaraayanan, 2023), while Acharya, Parlatore, and Sundaresan (2025) and Huang, Pagano, and Panizza (2020) study the interaction between public and private players. Garrett and Shive (2022) specifically study how U.S. federal tax incentives to investors in renewable power plants distort asset performance.

Moreover, a large literature in economics has studied policies that provide tax incentives to local investments with the goal of accelerating local economic development and job growth (see, for example, Greenstone, Hornbeck, and Moretti, 2010, Moretti and Wilson, 2014, and Kline and Moretti, 2014a, and the literature review by Bartik, 2018), and the effects of state taxation on business development (see recent contributions by Gale and Rueben, 2015, Suarez Serrato and Zidar, 2016, and Giroud and Rauh, 2019).

We provide novel findings on how local tax incentives shape private investment in data centers, and document the lack of significant spillovers on tech job creation.

Our paper also contributes to a growing body of work on the real effects of data center development. Benetton, Compiani, and Morse (2024) examine the impact of crypto mining on electricity prices. In contemporaneous work, Feher, Garcia-Appendini, and Mihet (2025) study more broadly the effects of local data center development on utility costs. Lee and Schmalz (2026) develop a theoretical model showing how negative externalities from elevated electricity costs affect welfare in the presence of financially constrained households. Brandsaas et al. (2025) propose a methodology for producing forward-looking estimates of data center investment for 2026 and 2027.

Finally, our work is related to the rapidly growing literature studying the effects of LLMs

and AI on a variety of outcomes.[7] The papers most closely related to our analysis focus on the impact of AI on workers' productivity (Brynjolfsoon, Li, and Raymong, 2025) and job creation. Over the past decade, demand for AI skills has surged (Alekseeva et al., 2021), accompanied by rapid job growth in the sector (Babina et al., 2024a), significant workforce reorganization (Babina et al., 2024b), and (so far) limited broader effects on the labor market (Acemoglu et al., 2022).

The paper proceeds as follows. Section 2 discussed the datasets used in our analysis, the institutional background of the data center sector, and the tax incentives provided by U.S. states. Section 3 provides the causal estimates of the effects of tax incentive laws on data center development, and discusses how the design of the statutes drives heterogeneous treatment effects. Section 4 estimates the operating cost savings generated by tax incentives, develops a quantitative model of hyperscale data center development decisions, and conducts counterfactual exercises in the absence of tax incentives. Section 5 discusses the evidence of the effects of data centers on local job markets. Section 6 concludes.

2 Background and Data

In this section, we describe our main datasets and present summary statistics (subsection 2.1), overview the background of the data center sector (subsection 2.2), and discuss tax incentive regulations (subsection 2.3).

2.1 Data

We obtain high-quality data on data center geolocations, years of construction, ownership, transaction history, and capacity. The data are made available by S&P Global. They cover hyperscale, wholesale, and retail data centers, and also contain information on facilities that

[7] Previous work has explored the influence of these technologies in numerous settings, including asset management (Sheng et al., (forthcoming), credit markets (Fuster et al., 2022; Choi et al., 2025), corporate governance (Cao et al., 2024), algorithmic pricing (Calvano et al., 2020), privacy and political control (Beraja et al., 2023).

are no longer operational as data centers, and on construction in progress.[8]

Table 1 displays summary statistics for data centers belonging to each category: hyperscale (panel A), wholesale (panel B), and retail (panel C). Hyperscale data centers have the largest scale in terms of power supply, square footage, and number of server racks. Specifically, the average hyperscale data center has a power capacity of 33 megawatt, covers an area of 168,000 square feet, and contains approximately 4,800 racks. Retail data centers are the smallest category, with a power capacity of 3,900 kilowatt, occupy 27,000 square feet, and include around 785 racks. Usage also varies substantially across categories: about 95% of hyperscale data centers have a single user, compared to only 4% among retail data centers. Finally, in terms of geographic distribution, hyperscale data centers tend to be located in areas with lower population density.

We complement the information on data centers with local employment data constructed using the dataset from RevelioLab, a workforce intelligence provider that compiles standardized individual-level microdata based on online professional profiles.[9] Although reliance on online professional profiles could raise concerns about sample selection, our analysis focuses on technology-sector workers, who are likely to engage with online tools and maintain online profiles and social media presence. The data provide information on both current occupations and employment history. For each individual, we observe the start and end dates of each job, the self-reported role, a description of the role (text from the online profile), the employer's name, and the job location (metropolitan area and city). The dataset also includes standardized fields that map individuals' occupational roles into standardized roles and industries.

We use these data to measure the number of new tech job starts across census places in

[8]The data are collected using a multi-source approach combining desk-based research, document analysis, direct stakeholder engagement, and field observations. Specifically, the data provider uses desk research to systematically review provider websites, construction announcements, and press releases. It then complements this information by examining municipal permitting documents, to verify project details. Additional information on data centers is also obtained through direct communication with other data vendors, including structured interviews and follow-up calls. Finally, site visits are conducted to validate selected observations and gather supplementary qualitative insights.

[9]Other papers using RevelioLab data include Cai et al. (2024), Canayaz, Erel, and Gurun (2024), Jin, Kermani, and McQuade (2025), and Eisfeldt et al. (2025).

states that are major data center markets (see Section 5). Summary statistics on job starts by location and sector from 2013 to 2025 for Texas, Virginia, and Ohio are presented in Internet Appendix Figure A.2.

2.2 The Data Center Sector

The data center sector represents a specialized asset class, distinct from other infrastructure assets and industrial real estate. The sector has experienced fast growth over the last 5 years, due to progress in cloud computing and Large Language Models, which require a massive computation infrastructure to operate at scale. However, the data center sector has existed for several decades, at least since the introduction of the internet.

Data centers are warehouses designed to host clusters of computer servers. They are equipped with advanced cooling systems, electrical systems, power back-up units, and redundant connections. This ensemble of equipment and structures allows data centers to keep servers functional effectively non-stop, even when facing local system failures or electricity outages.

Data centers are either owned and operated by IT, telecom, or technology companies, or alternatively owned by investors, that lease out their capacity to tenants' servers. Specifically, the investor owns the warehouse and related infrastructure, but not the servers, which are instead owned and installed by the tenants. Rents are not quoted as dollars per occupied square footage, but rather as dollars per amount of uninterruptible power supply (UPS) available to the tenants' servers (measured in kilowatt). The tenant typically pays rent plus reimbursements of electricity costs to the owner of the facility.

As mentioned in the previous section, there are three main categories of data centers in our dataset: hyperscale, wholesale, and retail. The original boom of internet technologies in the early 2000s led to the initial large-scale development of retail data centers, followed by a wave of wholesale facilities devoted to cloud computing, as this technology progressively became mainstream. In the last decade, most new facilities are large data centers. In particular, the

last few years have seen large growth in the development of hyperscale data centers. This has coincided with a large increase in the power demanded by the data center sector as a whole. Industry research[10] shows that in 2016, the entire data center sector in the U.S. had new leased and pre-leased capacity below 100 megawatt per quarter. In 2020, new leases and pre-leases were below 200 megawatt. In 2024 they exceeded 1,000 megawatt.

The technological waves described above are also reflected in the data we use for our study. Figure 1 plots the number of data center facilities completed in each year from 1994 to 2025. There are three time periods in which data center development peaks. The first peak is in the late 1990s and early 2000s, when the dotcom boom was leading to the expansion of network connections and the internet infrastructure. The second peak is in the early 2010s, and the last one is in the 2020s, corresponding to the introduction of LLMs and AI.

The different colors in the figure depict the number of retail (in blue), wholesale (in green), and hyperscale (in red) data centers. Retail and wholesale data centers are the most frequently developed assets until the 2020s. During the 2010s, we see a progressive increase in the number of hyperscale facilities completed each year. In the 2020s, almost all new facilities are hyperscale or wholesale.

Appendix Figure A.3 displays the number of operational, under construction, or planned data centers across states for each data center category. Hyperscale data centers have higher geographic concentration than retail data centers.

2.3 State Tax Incentives

U.S. states have implemented tax incentive schemes to stimulate the development of data centers at different points in time over the last two decades. We provide a summary of these policies in Table 2.

The first two columns list the state enacting a new tax incentive law and the year in which

[10] Information on the data center sector and its growth over time is available from the report by Guarino, Samandyk, and Webster (2025).

the policy was implemented. As of November 2025, 38 states have data center incentives in place, and no state has repealed such legislation. However, some states are actively debating whether to suspend their local incentive programs. In Georgia a bill to suspend the incentives passed the legislature but was vetoed by the governor in May 2024,[11] while the governor of Illinois has called in February 2026 for a two-year suspension of local tax incentives.[12]. Some states have revised incentives over time, typically to broaden eligibility criteria or to lower minimum investment thresholds.

New York enacted the first incentive scheme in the country in year 2000, while Kansas, West Virginia, and Massachusetts have enacted their policies most recently, with provisions taking effect in 2025. However, 20 out of the 38 states implementing tax incentive policies have introduced incentives schemes between 2009 and 2015 (see also Internet Appendix Figure A.1). This period significantly predates the most recent wave of data center development, and instead coincides with the period of rapid diffusion of pre-LLM cloud computing, which was driven by the commercial launch of major providers such as Amazon Web Services (2006), Microsoft Azure (2010), and Google Cloud Platform (2011).

Table 2 also provides a schematic overview of the main features of the tax incentives. Almost all states exempt data centers from sales and use taxes, while a smaller set of states also provides exemptions from property taxes. Incentives coverage generally includes tangible equipment (e.g., servers, data storage devices, and network connectivity equipment), and infrastructure components (e.g., including electrical systems, UPS, and cooling systems). Some states also exempt the electricity consumed in data center operations. Overall, all incentive programs impact data center performance by lowering operating costs or maintenance and capital expenditures. Thus, the incentives are designed to indirectly stimulate investment and development of data centers by offering higher margins to operators, and increasing asset values.

Nearly all states condition eligibility upon a minimum capital investment. As we discuss in

[11] For more details, see this newspaper article.
[12] See this article.

Section 3.4, this has important implications for the effects of these policies. Some states employ tiered systems that differentiate thresholds by population size or geographic location, while others establish a single statewide standard. Alabama and Connecticut impose the highest thresholds ($400 million), while Nebraska imposes the lowest ($1 million). When the tax benefits are available only for a limited time period (typically between 10 and 30 years) after development, the length of this period is strongly correlated with the minimum level of capital investment. States imposing a higher investment threshold offer incentives for a longer time.

In many states, eligibility also depends on the creation of a minimum number of full-time, well-compensated positions. This is consistent with regulators pursuing the goal of creating local economic development through subsidized data center facilities. However, it is important to note that the job-creation thresholds are minimal relative to the investment-scale requirements.

3 State Incentives and Data Center Development

In this section, we estimate the effect of state-level tax incentives on the development of new data centers (subsection 3.1), we then explore the relation between regulation scope and effect size (subsection 3.2), estimate the effects of incentives on asset-level characteristics (subsection 3.3), and discuss the mechanism driving heterogeneous responses to incentives across data center categories (subsection 3.4).

3.1 State-Level Evidence

To study whether local incentives have statistically and economically significant effects on the development of new data centers, and whether these effects differ substantially across data center categories, we estimate a staggered difference-in-differences design. We begin with a standard specification on a state-by-year panel:

$$y^k_{s,t} = \delta D_{s,t} + BX_{s,t} + a_s + a_t + e_{s,t} \tag{1}$$

where s is a U.S. state, t is a year (ranging between 1994 and 2025), $D_{s,t}$ is a dummy equal to one in states that introduce data center tax incentives and for years t greater or equal that the year in which the incentives are introduced. Finally, a_s and a_t are state and year fixed effects, and $X_{s,t}$ is a vector of state-level time-varying controls. The inclusion of controls can account for potentially confounding effects taking place at the same time as the laws, but can also bias estimates of the coefficient δ if the evolution of the control variables is also impacted by the introduction of state incentives. Thus, we report estimates of δ both from specifications that do not, and do include the vector of controls $X_{s,t}$.

The dependent variable $y^k_{s,t}$ measures the scale of new data center development for facilities belonging to category k (either hyperscale, wholesale, or retail), and completed in state s and year t. It is either a dummy equal to one when there is at least one data center facility completed of category k in state s and year t, or the log of one plus the number of completed data centers, or the number of completed data centers (in this case we estimate a Poisson regression, following Cohn, Liu, and Wardlaw, 2022), or the log of one plus the total UPS power capacity of completed data centers.

The coefficient of interest is δ, which has been interpreted in finance and economics studies as the estimate of the average treatment effect from staggered treatment. However, a recent literature has shown that the coefficient instead converges to the variance-weighted treatment effect for the treated, and only under restrictive assumptions (Baker, Larcker, and Wang, 2022). As a result, regression specifications of this type may provide biased estimates. To address this concern, we also report results based on the approaches developed by Cengiz et al. (2019), Sun and Abraham (2021), and Callaway and Sant'Anna (2021). Our estimates are robust across all these specifications.

Table 3 reports the baseline estimates. The vector of controls $X_{s,t}$ includes variables tracking electricity cost and consumption in state s: the lagged (in year $t-1$) log electricity price per kW for residential use, the log price per kW for business uses, and the lagged (between year $t-2$

and year $t-1$) increase in electricity consumption (in MW) for residential and business use.[13] Electricity is the largest operating expense for data centers, particularly for recent hyperscale and wholesale facilities. Strikingly, the estimates of the coefficient δ are virtually unaffected by the inclusion of these controls.

The introduction of tax incentives is associated with a significant increase in the completion of hyperscale data centers. The likelihood that at least one hyperscale facility is completed in a given year is 26 percentage points higher, while the number of completed data centers is between 26% (OLS estimate) and 350% (Poisson estimate) higher.[14] Total UPS power capacity completed is more than 250% larger. We find significant but weaker effects for wholesale data centers, with an 8 percentage point (marginally significant) increase in the likelihood of seeing at least one new project completed in a year, noisy effects on the increase in the number of completed data centers, and a significant 85% increase in UPS power. The effects on retail data centers are not statistically significant, with smaller point estimates than for the other two categories.

The results are similar when we use alternative estimators that address potential biases from the specification in equation 1. In this case, the dependent variables are the log number of completed data centers and the log power capacity of the completed projects. We report estimates in Table 4. The stacked regressions approach from Cengiz et al. (2019) and the cohort-specific relative-time treatment effects regression from Sun and Abraham (2021) produce estimates that are virtually identical to those from the naive regressions. The approach from Callaway and Sant'Anna (2021) produces similar results for hyperscale data centers, but not statistically significant estimates for wholesale, and even negative and significant estimates for retail data centers.

Figure 2 provides estimates of the dynamic treatment effects coefficients from the approach

[13] These data are available from the U.S. Energy Information Administration for the years from 1990 to 2023 (https://www.eia.gov/opendata/).

[14] The estimation algorithm for the Poisson regressions selects a smaller sample due to separation issues with states that have no data center projects completed over the entire time period of our study. For this specification, we estimate the semi-elasticity to the implementation of a law as $\exp(\delta) - 1 \approx 3.5$.

in Sun and Abraham (2021), with lead and lag coefficients covering the sample from 7 years before to 15 years after policies are implemented. The different panels display results for hyperscale, wholesale, and retail data centers. For the panels on the first row, the dependent variable is the log number of completed facilities, while for the panels on the bottom it is the log of UPS electric power capacity completed. The graphical evidence confirms the results displayed in the regression tables. There is no evidence of a treatment effect for retail data centers, while the strongest response is from hyperscale data centers. This effect becomes significant 3-4 years after the implementation of tax incentives, consistent with a development-completion lag.

In addition to the evidence on staggered treatment effects just discussed, we also find that the presence of state-level tax incentives in 2025 is associated with a more active pipeline of planned data center construction. Descriptive state-level regressions reported in Appendix Table A.1 show that the volume of planned hyperscale data center facilities is strongly associated with the presence of tax incentives. Results for the other data center categories are noisy.

3.2 Heterogeneous Effects from Regulation Scope

We test whether the design of tax incentive provisions drives heterogeneous responses across states. We construct for each state a measure called $IncentivesScope$, which captures how broad and significant are the incentives offered by the local law. For instance, a regulation that imposes a lower minimum investment size threshold is broader. A law that grants sales and use tax exemptions for both equipment and electricity is broader. A law that grants incentives for a longer time period after facility completion provides more significant benefits.

The construction of the index is complicated by the fact that the laws differ along multiple dimensions, and there are several degrees of subjectivity in how to evaluate differences between statutes and provisions. To limit the influence of researcher subjectivity, we construct the score using commercially available LLM algorithms to assign scores (see e.g., Bartik, Gupta, and Milo, 2025 and Bombardini, Trebbi, and Zhang, 2025). In order to attenuate concerns with

idiosyncratic features and biases of each individual algorithm (Liu, Liu, and Wang, 2025), we submit the same set of prompts to and obtain scores from four different commercially available platforms: ChatGPT, Claude, Grok, and Gemini. We ask each algorithm to provide a score from 1 to 5 for each law, along with a description of the motivations determining each score. The details of this process are described in Internet Appendix B. We find substantial overlap in scores and motivations across the board. Finally, we construct the index $IncentivesScope_s$ for each state s as the average score across the four algorithms. The values of the index for each state (along with the scores attributed by individual LLMs) are displayed in Internet Appendix Figure A.4.

We then estimate the following regression equation:

$$y^k_{s,t} = \gamma\big(D_{s,t} \times IncentivesScope_s\big) + BX_{s,t} + a_s + a_t + v_{s,t} \tag{2}$$

where $D_{s,t}$ is a dummy equal to one in states that introduce data center tax incentives and in years t greater or equal that the year in which the incentives are introduced, a_s and a_t are state and year fixed effects, and $X_{s,t}$ is a vector of state time-varying controls. Table 5 reports estimates of the coefficient of interest, γ, which captures the sensitivity of the increase in data center completions to the scope of local incentives.

Consistent with the results in Table 3 and Table 4, we find highly significant effects for hyperscale, significant but weaker results for wholesale, and overall insignificant effects for retail. Thus, the scope and significance of local incentives impact the response of data center development, which is in line with the causal interpretation of the evidence in the previous section. However, we still find significant heterogeneity across data center types, with hyperscale data centers being the most responsive.

3.3 Asset-Level Characteristics

We dig further into the effects of the tax incentives by exploring their impact on the characteristics of individual data centers developed across states. We find significant effects, but entirely explained by the fact that local incentives change the composition of data centers across categories. Once controlling for data center categories (hyperscale, wholesale, and retail), the effects of incentives on data center characteristics disappear.

We begin by estimating the effects of incentives with the following regression specification:

$$q_{i,t} = \phi D_{s,t} + BX_{s,t} + a_s + a_t + v_{i,t} \tag{3}$$

where $q_{i,t}$ is a characteristic of data center i built in year t, and all other variables have the same interpretation as in equation 1. The coefficient of interest is ϕ, which measures how characteristics change for newly built data centers, within the same state, after the implementation of local tax incentives.

First, we test whether incentives impact the physical and technological attributes of the data center. These are log size (measured as UPS power capacity), a dummy equal to one for data centers that have a single user, a dummy equal to one for data centers that are devoted to cloud services (rather than network or co-location), the log of power capacity per square foot (capturing the density of servers within the data center), and power usage effectiveness (PUE), which is the ratio of total power coming into the facility from the electricity provider over power deployed to the servers (UPS power). This last measure indicates how efficiently the data center uses electric power.

Second, we test whether incentives change the likelihood of having employees on site. We measure this using dummies equal to one for data centers that include offices within their main structure or have on-site maintenance staff.

Third, we test if incentives change the micro-location of data centers. In particular, we test whether after the passage of incentives, data centers are more likely to be located in remote

areas, identified by log of population density in the census tract in which the data center is built, and a dummy equal to one if the tract has low population density (we define this as population density below 50 individuals per square mile, which is approximately half of the median density across the U.S.).

Estimates of equation 3 are reported in Panel A of Table 6. We find that the introduction of tax incentives increases log size of power centers (in terms of UPS power) by more than 60%. It also increases the likelihood that a data center has a single tenant by 20 percentage points, and that it is devoted to cloud computing (either for traditional services or LLMs) by 16 percentage points. There is a marginally significant negative effect on employees on site, due to a 7 percentage point lower likelihood of having offices in the facility. Finally, there is a marginally significant effect on data center locations. New facilities are 5 percentage points more likely to be located in a low-density tract. We do not find significant effects for the remaining variables.

In summary, tax incentives lead to larger data centers, which are more likely to have a single user and to be devote to cloud service, marginally less likely to host offices, and marginally more likely to be located in rural areas.

In panel B, we expand the regression specification by including fixed effects for data center categories: hyperscale, wholesale, and retail. The point estimates for almost all effects shrink by two-thirds, and almost all coefficients become insignificant. Only the coefficient for the effect on log size remains borderline significant.

Thus, the effects are explained by the fact that incentives change the composition of new data centers across categories. Hyperscale data centers are larger, usually have a single user, and are used for cloud (LLM) computing. Because of their larger power and infrastructure requirements, they are slightly less likely to host offices and more likely to be built farther from high-population locations.

Figure 3 also shows that all hyperscale facilities in our data are owned by technology firms. Combined with the fact that these data centers have a single user, this suggests that these

facilities are owner-operated. The figure also shows that wholesale and retail data centers are instead frequently owned by real estate investors or smaller data center companies (typically LLCs owned by investors), which then rent out the electric power capacity to users that install their own servers. Thus, the tax incentive laws also increase the share of facilities that are owner-operated.

3.4 Tax Incentives and Data Center Scale

The evidence in the previous sections shows that state incentives have been highly effective in attracting the development of hyperscale data centers. Their impact on wholesale facilities is more limited, and they have virtually no effect on retail data centers. Through these heterogeneous effects across categories, the incentives have reshaped the local data center landscape.

The incentives impact several aspects of data center operations and capital expenditures, and it is at first unclear what are the main drivers of these sharp differences across categories.

We argue in this section that the crucial element is likely the fact that the laws impose substantially large minimum investment sizes. The thresholds are based either on a dollar investment amount or a minimum size. We find that in many circumstances these minimum requirements automatically exclude the scale at which retail data centers operate. They also exclude a substantial share of wholesale assets, but are instead a reasonable benchmark for hyperscale investments.

Using the information in Table 2, and estimates of data center development costs available from industry sources (Guarino, Samandyk, and Webster, 2025), we construct for each state the minimum square feet size that a new facility must meet to obtain the incentives. We explain the details of our calculations in Internet Appendix C.

To establish how binding these thresholds are for development decisions, we compare them against the distribution of facility sizes across the U.S. for the three categories of data centers. Then, for each state, we calculate what fraction of the operational stock of data centers in

the U.S. would meet the minimum threshold. The lower the share, the more it is likely that to meet the state-imposed minimum requirement developers may need to design a data center substantially larger than what is standard for that category.

Figure 4 presents box plots showing the distribution of the shares of qualifying data centers across state tax-incentive laws. In panel (a), the shares are calculated based on facility sizes in the U.S. as of 2020. We choose this year because at that point most of the state incentives programs had been implemented (see Table 2 and Figure A.1). However, the large growth in demand for large-scale hyperscale data centers, driven by LLMs workloads, had not yet occurred. Even at this point in time, comparing existing facility sizes to state-level thresholds makes clear that minimum investment requirements are far more compatible with hyperscale development. The median tax incentive provision across states would have qualified roughly 85% of the hyperscale stock, with an interquartile range across state provisions of 51% to 93%. Although wholesale facilities exhibit substantial variation in size, about 66% would have qualified for the median provision, with an interquartile range of 37% to 89%. By contrast, retail data centers typically fall below the required thresholds: in 2020, the median provision would have qualified 21% of retail facilities, with an interquartile range of 9% to 48%.

Repeating the same exercise using the stock of data centers in 2025 (panel b) gives even starker results. The size distributions for wholesale and retail facilities remain largely unchanged, and the distribution of the share of qualifying centers is similar to that observed in 2020. In contrast, the interquartile range for the share of qualifying hyperscale data centers compresses substantially, spanning only 66% to 95%.

The upward shift in the size distribution of hyperscale facilities reflects rising computational demands, but it may also be endogenously amplified by the investment thresholds in tax-incentive regulations. At the same time, because hyperscale facilities are inherently larger, they are better positioned to take advantage of the provisions, and do not need to distort their operational scale and technology to meet the requirements. As noted above, there is little change over time in the share of qualifying assets among smaller data centers. These facilities

rely on different technologies, and scaling up to meet the thresholds is not feasible. This is particularly true for retail data centers, which operate at scales far below the requirements.

4 Implications for Data Center Location Across States

This section focuses on quantifying the effects of tax incentives on data center development decisions across states. We first conduct back-of-envelope calculations to assess the magnitude of the subsidies delivered by state tax incentives (subsection 4.1). We then develop and calibrate a simple quantitative model that formalizes development decisions across states (subsections 4.2, 4.3, and 4.4). Finally, we use the model to develop counterfactual exercises showing how the distribution of new development projects is strongly influenced by tax incentives (subsection 4.5).

4.1 Stylized Facts on the Value of Incentives

In this section, we combine the statutory provisions of each law with information on electricity costs, sales taxes, and operating and capital expenditures to estimate tax savings per unit of scale, measured in UPS power.

For the facilities with a high likelihood of having qualified for incentives,[15] we construct estimates of tax savings using 2025 operational data. In this exercise, we focus on savings arising from exemptions from sales and use taxes. This is motivated both by the fact that these exemptions constitute the primary incentive mechanism across states, and by the fact that they are particularly relevant for data center operations. Electricity is the largest operating expense, and capital expenditures for replacing electrical equipment and infrastructure (e.g.,

[15]Figure A.5 in the Appendix reports, for each data center category, the share of facilities in our dataset built between 2000 (the year in which the first tax-incentive program was introduced) and 2025 that likely qualified for incentives. We provide both upper- and lower-bound estimates based on the range of thresholds specified under different conditions in state incentive laws. For hyperscale facilities, at least 60% are highly likely to have qualified. For wholesale facilities, approximately 40% appear to have qualified. In contrast, only 9% of retail facilities meet the criteria. These results are consistent with the evidence in Section 3.4.

cooling systems) can amount to 25% of operating income.

We begin by using the information in Table 2 to determine whether a facility is eligible for exemptions on: (i) sales and use taxes on electricity, (ii) sales and use taxes on equipment, and (iii) sales and use taxes on infrastructure. For each state, we then collect the corresponding sales tax rates for these three categories

Second, we construct estimates of electric-utility costs and capital expenditures for each qualifying data center. Electricity cost estimates are based on data from the U.S. Energy Information Administration on industrial electricity prices per kW by state and year, combined with consumption estimates based on each facility's UPS capacity and its power usage effectiveness (PUE). Capital expenditure estimates are derived from data center size and UPS power per square foot (see Appendix C for details).

Using these estimates, we compute for each data center the total 2025 sales tax savings for projected electricity consumption, equipment purchases, and infrastructure expenditures. Figure 5 presents a state-level heatmap of the average subsidy per megawatt of UPS capacity, calculated for data centers in our dataset that meet state-law requirements and are completed by 2025. States shown in white have active incentive programs but no qualifying data centers, while states shown in grey do not have incentive programs. Alaska and Hawaii are omitted because neither state has an incentive program.

Although Virginia hosts the largest number of data centers in the country, it does not provide the highest subsidy per unit of power, offering only about $22,000 per UPS MW. Indiana, Minnesota, and Tennessee provide the most generous incentives, approximately equal to $115,000, $100,000, and $95,000 per UPS MW, respectively. Texas, Iowa, and South Carolina also offer substantial incentives to existing qualifying data centers in 2025. Each state exceeds $80,000 per UPS MW. The mean operational hyperscale data center in our data has total UPS power of 33 megawatt. Thus, with incentives of $80,000 per UPS MW, the data center receives on average in 2025 savings on annual operating expenses of $2.64 million.

In commercial real estate, it is standard practice to use multiples of operating income

relative to asset prices to obtain quick, back-of-the-envelope valuations. These multiples are known as capitalization rates, or cap rates. For example, an asset that generates $100,000 in net operating income and is located in a market where the prevailing cap rate is 10% would be valued at $1 million ($100,000 / 0.10). Drawing on industry sources (Guarino, Samandyk, and Webster, 2025), we estimate that current cap rates for hyperscale data centers in the U.S. are on average approximately 6.5%.

We use the cap rate to convert the flow of operating cost savings from tax incentives into an implied asset value. For the mean hyperscale data center, the incentives translate into an additional $40.6 million in asset value ($2.64 million / 6.5%). Industry estimates indicate that developing a data center costs roughly $11 million per megawatt, implying a total development cost of $363 million for a 33 MW facility. Thus, the incentive value amounts to approximately 11% of development cost (40.6/363), a magnitude that is economically significant and is likely to impact development decisions.

4.2 Quantitative Model Outline

We develop a simple partial equilibrium model from the perspective of an investor deciding whether, and in which U.S. state, to build a hyperscale data center. The model predicts the spatial distribution of investment across states, incorporating state-level tax incentives and the option to delay investment. We draw on two established literatures: discrete choice models of the spatial location decisions of workers and firms (e.g., Kline and Moretti, 2014b, Suárez Serrato and Zidar, 2016, and Moretti and Wilson, 2017) and models of investment under uncertainty in the presence of fixed costs (Dixit and Pindyck, 1994).

Our model is kept deliberately simple. We abstract from endogenous responses in both the demand for data center services and the supply of development capital. These forces are shaped by global trends that extend beyond the within-U.S., cross-state setting on which we focus. We believe our framework captures some of these effects in reduced form through the dynamic state

variables, expected growth in operating income, and discount rates that we calibrate to current market conditions in the next section. Additionally, the option value of waiting provides insight into the incentives for capital to remain on the sidelines.

There are J possible development locations, corresponding to different U.S. states, indexed by $j \in \{1, \ldots, J\}$. Each location has a K-dimensional state vector $\mathbf{s}_{j,t}$, with vector autoregression (VAR) dynamics:

$$\mathbf{s}_{j,t+1} = \mathbf{c}_j + \mathbf{M}_j \mathbf{s}_{j,t} + \boldsymbol{\varepsilon}_{j,t+1}, \tag{4}$$

where $\mathbf{c}_j$ is a constant vector, $\mathbf{M}_j$ is a $K \times K$ VAR matrix for location j, and $\boldsymbol{\varepsilon}_{J,t+1} \sim N(\mathbf{0}, \boldsymbol{\Sigma}_j)$ is the residuals vector.

Operating income in location j at time t, net of costs and capital expenditures, is given by:

$$\pi_{j,t} = \sum_{k=1}^{K} B_{j,k} \cdot \exp(s_{j,t}^k) + \mu_j, \tag{5}$$

where the coefficients B_{ik} can be positive or negative, and determine how state variables influence income, while μ_j is a state-specific constant. The present value from investing in state j is equal to:

$$PV_j(\mathbf{s}_{j,t}) = E_t \left[\sum_{\tau=0}^{\infty} \beta_d^\tau \pi_{j,t+\tau} \right] = \sum_{k=1}^{K} B_{j,k} E_t \left[\sum_{\tau=0}^{\infty} \beta_d^\tau \exp(s_{j,t+\tau}^k) \right] + \frac{\mu_j}{1-\beta_d},$$

where β_d can be decomposed as follows: $\beta_d = \tilde{\beta}_d \exp(\delta)$; $\tilde{\beta}_d$ is a discount factor reflecting the projects' cost of capital and δ is a time invariant expected growth rate of income.

The equation above shows that the present value is equal to the sum of K components, each driven by the expected evolution of each state variable, and a constant term driven by the state-specific constant. Substituting the state variable dynamics in the present value equation,

we find:

$$PV_j(\mathbf{s}_{j,t}) = \sum_{k=1}^{K} B_{j,k} \left\{ \sum_{\tau=0}^{\infty} \beta_d^{\tau} \exp\left(\mathbf{e}_k' \left[\mathbf{M}_j^{\tau} \mathbf{s}_{j,t} + \sum_{j=0}^{\tau-1} \mathbf{M}_j^{l} \mathbf{c}_j \right] + \frac{1}{2} \mathbf{e}_k' \mathbf{\Omega}_{j,\tau} \mathbf{e}_k \right) \right\} + \frac{\mu_j}{1-\beta_d}, \quad (6)$$

where $\mathbf{e}_k$ is the k-th unit vector and $\mathbf{\Omega}_{j,\tau} = \sum_{l=0}^{\tau-1} \mathbf{M}_j^{l} \mathbf{\Sigma}_j (\mathbf{M}_j^{l})'$.

The investment decision is irreversible, and the net present value obtained by an investor in state j at time t is:

$$V_j(\mathbf{s}_{j,t}) = PV_j(\mathbf{s}_{j,t}) - F_j + u_{j,t},$$

where F_j is the fixed development cost in state j, and $u_{j,t}$ is an idiosyncratic valuation shock, capturing random unpredictable factors influencing project values (extreme events, shocks impacting local labor markets or development sites, etc.). An investor can choose to invest in any of the J states, or to delay investment. The value of delaying for one period is:

$$W_t = \beta_d E_t \left[\max_{j \in \{1,\ldots,J\}} V_j(\mathbf{s}_{j,t+1}) \right]$$

We can then derive the probability of investing in each state, and the probability of delaying. Specifically, if valuation shocks follow a Type-I Extreme Value distribution (McFadden, 1978), we can write the probability of investing in state j at time t as:

$$P_{inv,j,t} = \frac{\exp(V_j(\mathbf{s}_{j,t}))}{\sum_{i=1}^{J} \exp(V_i(\mathbf{s}_{i,t})) + \exp(W_t)},$$

The probability of delaying for one period is:

$$P_{W,t} = \frac{\exp(W_t)}{\sum_{i=1}^{J} \exp(V_i(\mathbf{s}_{i,t})) + \exp(W_t)},$$

where $W_t \approx \beta_d \ln\left(\sum_{i=1}^{J} \exp\left(E_t[V_i(\mathbf{s}_{i,t+1})] \right) \right)$.

The probability of investing in state j, conditional on not waiting is:

$$P^{cond}_{inv,j,t} = \frac{P_{inv,j,t}}{1 - P_{W,t}}.. \tag{7}$$

This probability is the main outcome of interest in our analysis, because it can be mapped to the realized distribution of investments across states observed in the data.

4.3 Model Calibration

We calibrate the model described in the previous section to match predicted investment probabilities to the distribution of the hyperscale data center development pipeline across U.S. states in 2025. We do not explicitly model project scale decisions; accordingly, we calibrate the model to reflect magnitudes per UPS megawatt of hyperscale capacity. Calibrated values of key parameters are reported in Appendix Table A.2.

For the calibration, we consider investment decisions across $J = 30$ states, which had the highest share of operating data centers as of 2020, and host most of the development pipeline in 2025. The states included in our exercise are: Arizona, California, Colorado, Florida, Georgia, Illinois, Indiana, Iowa, Kansas, Kentucky, Maryland, Massachusetts, Michigan, Minnesota, Missouri, Nebraska, Nevada, New Jersey, New York, North Carolina, Ohio, Oregon, Pennsylvania, South Carolina, Tennessee, Texas, Utah, Virginia, Washington, and Wisconsin.

We obtain $\beta_d = 0.938$ by setting $\tilde{\beta}_d = 0.91$, which reflects the expected unlevered rate of return for operating data centers recommended by industry research, and $\delta = 3\%$, the projected growth rate of operating income.

The state vector $\mathbf{s}_{j,t}$ has three elements. The first is the log of state-level monthly electricity prices, available from the U.S. Energy Information Administration. The second and third element are indices, capturing the evolution of electric equipment and data center infrastructure costs. These two series are common across all states. The electrical equipment cost series is constructed as an index whose growth rate equals the average of the growth rates of the Pro-

ducer Price Indices (PPIs) for electrical equipment manufacturing (PCU33533353), other electrical equipment and component manufacturing (PCU33593359), switchgear and switchboard apparatus manufacturing (PCU335313335313), and electric power and specialty transformer manufacturing primary products (PCU335311335311P). The data center infrastructure cost series is an index whose growth rate equals the average of the PPIs for communication and energy wire and cable manufacturing (PCU3359233592) and HVAC and commercial refrigeration equipment (PCU33343334).

We then log and demean the series and estimate equation 4 for each state at monthly frequency over the period from January 2004 to July 2025. Since the series are demeaned, the intercept vector $\mathbf{c}_j$ is a vector of zeros. We adjust the mean reversion matrix $\mathbf{M}_j$ and the variance-covariance matrix of residuals $\mathbf{\Sigma}_j$ for each state to reflect annual dynamics.

In our calibration, we scale income in millions of dollars. We assume a baseline fixed investment cost per UPS megawatt of $C = \$11$ million for all states, consistent with the estimates discussed in the previous section. The construction cost includes equipment and infrastructure, and thus benefits from sales tax incentives. Accordingly, the fixed cost for state j is $F_j = C \times (1 - EqSalesTaxSh_j - InfrSalesTaxSh_j)$. The incentives are calculated as the product of the local sales tax rate in 2025 ($SalesTax_j$) and the share of cost attributable to electrical equipment and infrastructure, which we compute based on the assumptions discussed in Appendix B for the back-of-the-envelope calculations.

We then calibrate the coefficients $B_{j,k}$ to capture the effects of the state variables on revenues and costs as follows:

$$
\begin{aligned}
B_{j,1} &= f_{j,1} - \kappa_{j,1} = \delta_{j,1} - ElCost_j \times \left[1 + \left(ElSalesTax_j \times I_j^{NoElIncentive}\right)\right], \\
B_{j,2} &= f_{j,2} - \kappa_{j,2} = \delta_{j,2} - DeprElEq_j \times C \times \left[1 + \left(SalesTax_j \times I_j^{NoEqIncentive}\right)\right], \\
B_{j,3} &= f_{j,3} - \kappa_{j,3} = \delta_{j,3} - DeprInfr_j \times C \times \left[1 + \left(SalesTax_j \times I_j^{NoInfrIncentive}\right)\right],
\end{aligned}
$$

where the $f_{j,k}$ coefficients capture revenue effects and the $\kappa_{j,k}$ coefficients capture cost effects.

We begin by describing the calibration of $\kappa_{j,k}$. In the $B_{j,1}$ equation, $ElCost_j$ is the average annual industrial electricity cost per megawatt in state j and in 2025, while $ElSalesTax_j$ is the electricity sales tax for state j in year 2025, and $I_j^{NoElIncentive}$ is a dummy equal to one if the state does not offer electricity sales tax incentives in 2025. In the $B_{j,2}$ and $B_{j,3}$ equations, the coefficients $DeprElEq_j$ and $DeprInfr_j$ match CAPEX for equipment and infrastructure per megawatt as a share of investment cost, using estimates from the back-of-the-envelope calculations, which are discussed in detail in Appendix C. $SalesTax_j$ is the sales tax in state j and in 2025, while $I_j^{NoEqIncentive}$ and $I_j^{NoInfrIncentive}$ are dummies equal to one if state j does not offer incentives for equipment and infrastructure in 2025.

The dummies $I_j^{NoElIncentive}$, $I_j^{NoEqIncentive}$, and $I_j^{NoInfrIncentive}$ match the information displayed in Table 2, to reflect tax incentives as of 2025 across states.

Given the calibrated $\kappa_{j,k}$ components, the vectors $\mathbf{s}_{j,t}$ containing the current values of the time series variables in each state are normalized to vectors of zeros.

The coefficients $f_{j,k}$ capture revenue effects, reflecting the fact that fluctuations in the state variables $\mathbf{s}_{j,t}$ coincide with stronger demand for computing services and infrastructure, and thus higher revenues for data centers. We assume $f_1 = f_2 = f_3 = f$ for all states and set $f = \alpha C$, where $\alpha = 0.2$, which combined with our choice of β_d delivers reasonable valuations and cap rates, as we discuss in the next section.

Using the values of β_d, B_j, μ_j, and the time series dynamics of the state variables, the model delivers state-level present values (equation 6), net present values, the value of delaying by one year, and choice probabilities across states.

Given all other parameters, we estimate the constant state-specific component of operating income μ_j to align the probability of investing in state j, conditional on not delaying, with the share of under-construction and planned hyperscale data center projects in that state relative to all projects across the states included in the calibration. However, we limit these adjustment to a predetermined range. We impose $\mu_j \in [-0.18, 0.18]$, a range of approximately $\pm 20\%$ relative to the mean initial (2025) operating income per megawatt across states in the model (0.82).

4.4 Model Fit

In our data we track a total of 387 hyperscale projects under development in 2025, including both projects under construction and projects for which construction is planned in the near future. Figure 6 shows the fit between conditional investment probabilities and the share of new hyperscale projects across states.

The model replicates the investment allocation in the data, with investment probabilities matching project development shares in the main markets: Virginia, Ohio, Indiana, and Arizona. This is achieved generating q reasonable distribution of valuations. The mean present value of future operating income per megawatt across states is approximately $11.3 million, close to the fixed investment cost. Indeed, the mean net present value is $0.62 million. The ratio of operating income over data center valuations (π/PV) is 6.3%, which is slightly higher than industry estimates of cap rates net of CAPEX, which are between 5.5% and 6%.

Table A.3 provides a summary of model results across states, displaying present values of cash flows per megawatt, net present values, conditional investment probabilities (equation 7), and shares of the total number of projects under construction and planned in 2025. The probability of delaying investment by one year is 40%, which suggests that there might still be substantial capital waiting on the sidelines, and that the data centers pipeline may substantially increase in the next years, consistent with views expressed by industry experts and news reporters.[16]

The calibration of the fixed cost component μ_j allows us to capture factors that influence the spatial distribution of data centers but are not related to the time series factors included in the vector $\mathbf{s}_{j,t}$.[17] Table A.4 in the Appendix reports the results from regressions of the estimated μ_j on state-level characteristics. They show that μ_j tends to be higher in states with greater population and population density, higher median income, lower shares of workers

[16] See, for example, this article by McKinsey & Co.

[17] Indeed, setting $\mu_j = 0$ across all states generates similar average project valuations and cap rates but systematically shifts investment probabilities toward states with lower electricity costs.

in agriculture and manufacturing, and a higher share of individuals with bachelor's degrees. Interestingly, the values of μ_j are also negatively correlated with distance to one of the ten major fiber nodes in the country.[18] Thus, to match the data, μ_j produces an adjustment that benefits states with higher population and are richer, and with economies that are more dependent on services. These variables likely capture the fact that these states have better infrastructure and more developed local economies, which is going to be correlated with higher profitability of development projects.

4.5 Counterfactuals: Removing Tax Incentives

We use the model described in the previous sections to conduct counterfactual exercises, which investigate the effects of tax incentives on the spatial distribution of new hyperscale projects across states. In the first exercise, we maintain the calibration discussed in Section 4.3 but re-estimate the model J times, each time removing incentives for a single state j. We then assess how this affects the conditional development probability (P_{inv}^{cond}) for the state. Specifically, we compute $P_{invNI,j}^{cond}/P_{inv,j}^{cond} - 1$, where $P_{inv,j}^{cond}$ is the baseline conditional investment probability and $P_{invNI,j}^{cond}$ is the conditional investment probability when removing incentives only from state j.

Panel (a) of Figure 7 displays the results. We find large adjustments, consistent with our empirical results. Most states, were they to unilaterally terminate their incentives, would see a decrease in investment probability exceeding 50%, which we interpret as a corresponding contraction in their expected development pipeline. Some of the largest effects are for Indiana, Tennessee, Iowa, South Carolina, North Carolina, and Texas. These are states with generous

[18] We construct this measure collecting the coordinates of the following major U.S. fiber nodes: Ashburn/Sterling interconnection cluster (Loudoun County, VA), Herndon/Reston TYSONS corridor (Fairfax County, VA), 60 Hudson St, 111 8th Ave (New York County/Manhattan, NY), NJFX subsea/terrestrial meet (Monmouth County, NJ), 350 E Cermak (Digital Realty) and Equinix CH1 (Cook County, IL), One Wilshire (624 S Grand) and CoreSite LA1 (Los Angeles County, CA), Equinix MI1/NAP of the Americas (Miami-Dade County, FL), Dania Beach & Hollywood CLS (Broward County, FL), JAX NAP/Jacksonville CLS (Duval County, FL), Virginia Beach Cable Landing Station at Corporate Landing (Virginia Beach Independent City, VA), Flexential/Aligned and others in Hillsboro (Washington County, OR), Pacific City CLS serving NCP, Jupiter, and Hawaiki (Tillamook County, OR), multiple carrier hotels and cloud PoPs (King County, WA), Infomart/Equinix DA series and others (Dallas County, TX), and multiple campuses in Mesa/Phoenix (Maricopa County, AZ).

incentive packages, as discussed in Section 4.1. There are no effects for California and Oregon, as they do not have tax incentives in place. Oregon does not have state sales and use taxes.

We conduct a second counterfactual exercise in which we remove incentives from all states and compare the baseline investment distribution across states from Section 4.4 with the resulting counterfactual distribution. In the aggregate, removing incentives lowers the average net present value per megawatt across states to −$0.43 million and reduces present values in the states with the most generous incentives (Indiana, Iowa, Texas, North Carolina, South Carolina) by approximately $2 million, in line with the back-of-the-envelope calculations in Section 4.1. Interestingly, removing incentives only marginally changes the probability of waiting, increasing it from 40% to 42%. Thus, the main counterfactual effects are not on investment volume but on the allocation of investments across states. Table A.5 in the Appendix displays the distribution of present values, net present values, and investment probabilities across states.

Panel (b) of Figure 7 displays the difference between the counterfactual and baseline conditional investment probabilities for each state (ΔP_{inv}^{cond}). The changes are substantial, and the total absolute change in probabilities across states ($\sum_j |\Delta P_{inv,j}^{cond}|$) equals 23%, implying that in expectation more than 10% of hyperscale projects are reallocated.

The states experiencing the largest decreases are among those highlighted in Figure 5 as offering the most generous tax savings. Indiana experiences the largest drop (3.8%), followed by North Carolina (1.5%) and Iowa (1.2%). For these three states, data center development probabilities are more than halved under the counterfactual. Other states experiencing substantial declines include Texas, Ohio, Tennessee, and South Carolina. The state benefiting most from the reallocation is Oregon, which gains 5%. One reason Oregon benefits under the counterfactual is that it imposes no state-level sales or use taxes. Other states benefiting from the reallocation include Arizona, Georgia, and Ohio.

The exact magnitudes emerging from our exercise should be interpreted with caution, as our framework is relatively simplified, as mentioned above. Nevertheless, the results suggest that tax incentives may be driving a substantial reallocation of the data center pipeline across

U.S. states.

5 Data Centers and Local Labor Markets

In this section, we turn to the real effects of data centers on state economies. Although this is a broad question, we focus on labor market implications, as job creation is frequently cited by state and local governments as a primary objective of these incentive policies. For example, Indiana Governor Eric Holcomb, in discussing new data center development in the state, emphasized that: *"Indiana is committed to being a central hub in the global economy of the future, and this latest announcement ensures Hoosier communities and talent will be key to widespread advancements in cloud and artificial intelligence technology."*[19]

There is no doubt that hyperscale and large wholesale data centers are substantial development projects that generate temporary demand for construction workers and technicians. The key question, however, is whether they have lasting effects on local talent and technology jobs. For instance, are the tech companies that own or operate these facilities more likely to establish hubs or offices nearby?

We explore the relationship between data center development and local tech job creation, focusing on the states of Arizona, Iowa, Minnesota, North Carolina, Nevada, Ohio, South Carolina, Texas, and Virginia. We select these states because, based on our analysis in the previous section, they offer large incentives for new data center development. Moreover, among states that currently offer incentives, the states listed above have seen significant development of hyperscale and wholesale facilities in the last fifteen years.

We focus on the years 2013 to 2025, a period that saw the introduction of state incentive laws in many of the states in this sample and, as shown in Figure 1, a marked increase in the development of large data center facilities. This gives us enough observations to study how the development of large data centers affects job creation over time.

[19] See `www.iedc.in.gov/events/news/details/2024/06/04/gov-holcomb-announces-plans-for-new-1b-microsoft-data-center-in-northwest-indiana`.

We first describe how we construct measures of tech job starts (subsection 5.1), and then present evidence on the effects of data centers on local employment (subsection 5.2).

5.1 Measuring Tech Job Starts

We combine information on the micro-location of individual data centers with information on new tech employment, which we construct using data from RevelioLab. These data report information for millions of individuals who started at least one new occupation between 2013 and 2015. The data allow us to observe individuals' work histories, including roles, employers, and job descriptions. Additional details on the dataset are provided in Section 2.1.

Using these work histories, we construct annual estimates of new tech sector job starts for each census place (city, town, or unincorporated area) across all states included in our sample.

Because "tech worker" is not a standardized occupational category, and any categorization is subject to measurement error, we report results based on three alternative definitions. First, we use the job role labels provided by RevelioLab, which map individual occupations into categories of varying granularity, ranging from roughly 50 to 1,500 labels. For the main analysis, we use the mapping based on 150 labels,[20] even though our results are robust to using more granular categorizations.

Second, we construct a task-based measure (see, e.g., Alekseeva et al., 2021 and Acemoglu et al., 2022) that identifies tech workers by counting the occurrence of key technical terms in the job-duty descriptions included in their profiles.[21] This approach may overstate the number of tech workers, as self-reported job descriptions can offer a biased view of an individual's professional specialization.

Finally, we construct a broader measure, aimed at capturing not just tech workers, but all workers employed by tech companies. We use 6-digits NAIC codes (available in the RevelioLab

[20] We use the following labels to identify tech workers: Application Engineer, Automation Engineer, Data Analyst, Data Engineer, Data Scientist, Electrical Engineer, IT Analyst, IT Specialist, IT Project Manager, Information Security, Process Engineer, Software Developer, Software Engineer, and Technology Analyst

[21] Specifically, we flag profiles that mention any of the following keywords: Machine Learning, Deep Learning, Neural Network, Computer Vision, Artificial Intelligence, Software, Algorithm, Cloud, or Data Center.

dataset) to map individual employers into tech-related sectors. Appendix Table A.6 contains the list of codes we use to classify tech companies. As in Acemoglu et al. (2022), we use several codes from the 51 (Information) and 54 (Professional, Scientific, and Technical Services) sectors to capture service-related jobs. We also include codes from the 31 (Manufacturing) sector to account for potential effects that data centers may have on the computational hardware required to operate them. Although each of the three measures has limitations and may be subject to specific biases, we find that they are highly correlated, with Spearman correlation coefficients above 0.8.

Moreover, the data yield a sensible spatial distribution of tech jobs. Figure A.6 in the Internet Appendix plots an example for Texas. It displays total tech job starts between 2013 and 2025 by census place. We observe higher concentrations in the major urban areas, though not in proportion to population size. Austin exhibits by far the largest concentration of tech job starts, followed by the metropolitan areas of Houston and Dallas.

5.2 Data Centers and Tech Jobs

We investigate the relationship between data center development and tech employment, estimating the following OLS and Poisson regression equations:

$$log(N_{TechJobs,c,t+h} + 1) = \beta z_{DCenter,c,t} + a_c + a_t + e_{c,t}), \tag{8}$$

$$N_{TechJobs,c,t+h} = \exp(\psi z_{DCenter,c,t} + p_c + p_t + u_{c,t}), \tag{9}$$

where $N_{TechJobs,c,t+h}$ is a measure of the number of new tech jobs in census place c and year $t+h$, $z_{DCenter,c,t}$ is a variable measuring data center development in proximity of the same census place or city, in year t, and a_c and a_t (or p_c and p_t) are census place and year fixed effects. Including these fixed effects absorbs differences across locations in the scale of new tech job creation, as well as state-level trends in tech job growth.

We define $z_{DCenter,c,t}$ as a dummy equal to one if at least one wholesale or hyperscale data

center is completed in year t within 10 miles of the census place boundaries.[22] This accounts for the possibility that facilities situated in the vicinity of a city or town may still generate local job market spillovers. Standard errors are double-clustered by census place and year.

We focus on the two larger data center categories because retail facilities are not affected by the tax-incentive programs, as discussed in previous sections.

The coefficients of interest, β and ψ, capture the relationship between new data center completions and subsequent tech job creation. We estimate equation 8 and equation 9 separately for each state and for each of the three measures of new tech job starts described in the previous section. Our sample is restricted to the years from 2013 to 2025. As mentioned above, this is the period during which most tax incentive programs were introduced and when large-scale hyperscale data center development began to accelerate (see Figure 1).

Data center locations are not randomly distributed across space. Even when developers target low-density areas, they are likely to favor sites near more dynamic towns or urban areas, where access to labor, infrastructure, and other resources is easier. These locations may also be more likely to attract an increasing number of tech job starts over time, independent of spillovers from data centers. To mitigate these concerns, we construct our estimation sample using a matching algorithm.

Specifically, we retain in the estimation sample only (i) places that experience at least one data center completion (at least one occurrence of $z_{DCenter,c,t} = 1$) during our sample period, and (ii) matched places that never had a data center developed over the entire period.

For each exposed place, we match n (either one or three) non-exposed locations that are most similar in terms of demographic characteristics. The matching procedure relies on data from the 2012 five-year American Community Survey and uses the following characteristics: log median household income, log median home value, median age of household heads, log total population, poverty rate, and the share of the population with a bachelor's degree or

[22] In all states, conditionally on development taking place ($z_{DCenter,c,t} = 1$), the median and average number of developed data centers within the 10-mile buffer of the census place in a specific year is between one and two.

higher. This approach ensures that exposed and non-exposed locations are comparable along key socioeconomic dimensions prior to the start of our estimation period.

Table A.7 in the Appendix reports mean differences in the matched demographic variables between exposed places and matched non-exposed places. In most states, these differences are not statistically significant. The exceptions are Texas and Virginia, where exposed places exhibit significantly higher incomes and home values, lower poverty rates, and a larger share of household heads with a bachelor's degree or higher.

While matching does not fully eliminate the spatial selection of data center sites in these two states, it highlights that selection would, if anything, bias our estimates toward finding positive effects of development. Higher-income and more highly educated areas are generally more likely to attract tech job starts regardless of data center activity.

Figure 8 reports the estimated coefficients β (from equation 8, top panel) and ψ (from equation 9, bottom panel). We set the horizon h to one year, so the dependent variable captures the number of new tech jobs that begin in the year following the completion of a data center development. We measure new tech jobs using each of the three approaches discussed above. The β estimates are generally noisy and not statistically distinguishable from zero. The ψ estimates are somewhat more precise, but also mostly insignificant. Ohio is the only state with positive and significant coefficients across all three measures of job starts, with point estimates just below 10%.

One might argue that a one-year horizon is too short, and that labor market responses may take several years to unfold. Thus, in Figure 9 we repeat the analysis using as a dependent variable tech job starts three years after data center completion ($h = 3$). We again find no consistent positive evidence of excess tech job creation for almost all states. Ohio's ψ coefficients remain significant for two of the three job start measures, but their magnitudes decline, with point estimates ranging between 1% and 5%.

We explore whether our results are sensitive to sample construction criteria in Figures A.7 (for β estimates) and A.8 (for ψ estimates) in the Internet Appendix. We report results using

the full sample of census places and cities within each state, without matching, as well as results from our matching procedure with $n = 3$, which matches each exposed place with three non-exposed counterparts. Across these alternative samples, we find only minimal variation in point estimates or confidence intervals, both when considering job creation at a one-year and a three-year horizon.

Overall, there is no clear evidence of a meaningful association between data center development and local tech job creation. Most estimated effects are statistically insignificant, and significant point estimates appear to quickly decay over time, when comparing horizons of one and three years. For Ohio, which has the most robust estimates, our results imply only 1% to 5% additional job creation after three years. To put this in perspective, census places with a subsidized data center within 10 miles experience, on average, roughly 120 new tech job starts per year across all states, and about 150 in Ohio. Thus, the implied magnitudes correspond to only 2 to 7 additional jobs annually.

Based on our estimates in Section 4, we can calculate the annual subsidy for an average size 33 megawatt data center in Ohio, which is approximately $1 million. Combined with the implied job effects, this corresponds to a cost of roughly $140,000 to $500,000 per additional new job created each year. The trade-off is even less favorable in the other states we study, where job creation effects are frequently insignificant and subsidy levels are often higher than in Ohio. For example, although the estimates for Texas are not statistically significant, if we use them for a similar back-of-the-envelope calculation we find subsidies of $750,000 per job.

Thus, the results suggest that for state incentive programs to be worthwhile, their primary benefits must extend well beyond local tech job creation.

6 Conclusion

Data centers are a critical component of modern technological infrastructure and have attracted significant attention from U.S. state regulators, who have enacted tax incentive laws specifically

aimed at attracting data center development.

We study whether these laws have influenced state-level data center development, and through which mechanisms. We then quantify the value of incentives for data center developers and investors. Finally, we assess the extent to which local communities benefit through the creation of tech jobs.

Using state-of-the-art staggered difference-in-differences methods, we find significant effects only for large-scale data centers, specializing in high-intensity cloud computing and, more recently, LLM and AI workloads. We show that this pattern is likely driven by the design of the laws, which target large projects and therefore exclude facilities focused on network services or technologies that operate at smaller scales.

The incentives are structured to provide long-term operating cost savings, which increase asset values. We estimate that, in several states, annual operating cost savings can exceed $3 million for an average-sized hyperscale data center. We then develop a simple quantitative model to show how these savings deliver substantial value to investors and developers, and influence the location of data centers across U.S. states.

The justification for this long-term value creation by U.S. states should rest on the extent to which higher valuations are, at least in part, recaptured by local communities. One potential channel is the development of a durable base of high-paying tech jobs. Using granular data on individual job histories, we test whether data center development is associated with tech job creation. We find no meaningful or robust evidence that data center completions are followed by increases in new tech-job starts large enough to justify the size of the subsidies.

Our work is among the first to shed light on the complex economics surrounding subsidies for the tech sector's emerging infrastructure. Given the scale of these policies, it is crucial for regulators to identify the concrete benefits they deliver to local economic development.

References

Acemoglu, D., D. Autor, J. Hazell, and P. Restrepo. 2022. Artificial intelligence and jobs: Evidence from online vacancies. *Journal of Labor Economics* 40:830–50.

Acharya, V., C. Parlatore, and S. Sundaresan. 2025. Financing infrastructure in the shadow of expropriation. *Review of Financial Studies* 38:1468–18.

Agarwal, S., M. Ayyagari, Y. Cheng, and P. Ghosh. 2025. Road to stock market participation. Working Paper.

Agarwal, S., P. Ghosh, A. Mukhrjee, and S. L. Naaraayanan. 2023. Rural roads, small business creation, and the gender-gap in entrepreneurship. Working Paper.

Agarwal, S., A. Mukhrjee, and S. L. Naaraayanan. 2023. Roads and loans. *Review of Financial Studies* 36:1508–47.

Alekseeva, L., J. Azar, M. Ginè, S. Samila, and B. Taska. 2021. The demand for ai skills in the labor market. *Labour Economics* 71:1002–20.

Andonov, A., R. Kraussl, and J. Rauh. 2021. Institutional investors and infrastructure investing. *Review of Financial Studies* 34:3880–934.

Arefeva, A., M. A. Davis, A. C. Ghent, and M. Park. 2024. The effect of capital gains taxes on business creation and employment: The case of opportunity zones. *Management Science* 71:4566–81.

Babina, T., A. Fedyk, A. He, and J. Hodson. 2024a. Artificial intelligence, firm growth, and product innovation. *Journal of Financial Economics* 151:1037–55.

———. 2024b. Firm investments in artificial intelligence technologies and changes in workforce composition. Working Paper.

Baker, A., D. Larcker, and C. Wang. 2022. How much should we trust staggered difference-in-differences estimates? *Journal of Financial Economics* 144:370–95.

Bartik, A., A. Gupta, and D. Milo. 2025. The costs of housing regulation: Evidence from generative regulatory measurement. Working Paper.

Bartik, T. J. 2018. "but for" percentages for economic development incentives: What percentage estimates are plausible based on the research literature? Upjohn Institute Working Paper 18-289.

Benetton, M., G. Compiani, and A. Morse. 2024. When cryptomining comes to town: High electricity-use spillovers to the local economy. Working Paper.

Beraja, M., A. Kao, D. Y. Yang, and N. Yuchtman. 2023. Ai-tocracy. *Quarterly Journal of Economics* 138:1349–402.

Bombardini, M., F. Trebbi, and M. B. Zhang. 2025. Measuring the costs and benefits of regulation. *Annual Review of Economics* 17:345–65.

Boudot-Reddy, C., and A. Butler. 2024. Paving the road to re-election. *Journal of Public Economics* 239:10522–38.

Brandsaas, E. E., D. Garcia, R. Kurtzman, J. Nichols, and A. Zytek. 2025. Estimating aggregate data center investment with project-level data. Working Paper.

Brynjolfsoon, E., D. Li, and L. Raymong. 2025. Generative AI at work. *Quarterly Journal of Economics* 83:1–54.

Cai, W., Y. Chen, S. Rajgopal, and L. Azinovic-Yang. 2024. Diversity targets. *Review of Accounting Studies* 29:2158–208.

Callaway, B., and P. Sant'Anna. 2021. Difference-in-differences with multiple time periods. *Journal of Econometrics* 225:200–30.

Calvano, E., G. Calzolari, V. Denicolò, and S. Pastorello. 2020. Artificial intelligence, algorithmic pricing, and collusion. *American Economic Review* 110:3267–97.

Canayaz, M., I. Erel, and W. Y. Gurun, Umit G. 2024. When protectionism kills talent. Fisher College Working Paper.

Cao, S., W. Jiang, B. Yang, and A. L. Zhang. 2024. How to talk when a machine is listening: Corporate disclosure in the age of ai. *Review of Financial Studies* 36.

Cengiz, D., A. Dube, A. Lindner, and B. Zipperer. 2019. The effect of minimum wages on low-wage jobs. *Quarterly Journal of Economics* 134:1405–54.

Chen, J., E. Glaeser, and D. Wessel. 2023. The (non-)effect of opportunity zones on housing prices. *Journal of Urban Economics: Insight* 133:103451–.

Choi, J., D. Huang, Y. Zhishu, and Q. Zhang. 2025. How good is ai at twisting arms? experiments in debt collection. Working Paper.

Cohn, J. B., Z. Liu, and M. Wardlaw. 2022. Count (and count-like) data in finance. *Journal of Financial Economics* 146:529–51.

Corinth, K., and N. Feldman. 2024. Are opportunity zones an effective place-based policy? *Journal of Economic Perspectives* 38:113–36.

Dixit, A. K., and R. S. Pindyck. 1994. *Investment under uncertainty*. Princeton, NJ: Princeton University Press.

Eisfeldt, A. L., G. Shubert, M. B. Zhang, and B. Taska. 2025. Generative ai and firm values. *Journal of Finance (forthcoming)* .

Feher, A., E. Garcia-Appendini, and R. Mihet. 2025. Working Paper.

Fuster, A., G.-P. P., T. Ramadorai, and A. Walther. 2022. Predictably unequal? the effects of machine learning on credit markets. *Journal of Finance* 82:5–45.

Gale, William G, K. A., and K. Rueben. 2015. The relationship between taxes and growth at the state level: New evidence. *National Tax Journal* 68:919–41.

Garrett, D., and S. Shive. 2022. Power banks: Do tax equity investors add value to renewable power projects? Working Paper.

Giroud, X., and J. Rauh. 2019. State taxation and the reallocation of business activity: Evidence from establishment-level data. *Journal of Political Economy* 127:1262–316.

Greenstone, M., R. Hornbeck, and E. Moretti. 2010. Identifying agglomeration spillovers: Evidence from winners and losers of large plant openings. *Journal of Political Economy* 118:536–98.

Guarino, D., E. Samandyk, and N. Webster. 2025. Global data centers outlook. Tech. Rep. 1876R, Green Street.

Gupta, A., S. Van Nieuwerburgh, and C. Kontokosta. 2022. Take the q train: Value capture of public infrastructure projects. *Journal of Urban Economics* 129:1034–42.

Howell, S. T., Y. Jang, H. Kim, and M. S. Weisbach. 2023. All clear for takeoff: Evidence from airports on the effects of infrastructure privatization. Working Paper.

Huang, Y., M. Pagano, and U. Panizza. 2020. Local crowding-out in china. *Journal of Finance* 75:2855–98.

Jin, Z., A. Kermani, and T. McQuade. 2025. Native-immigrant entrepreneurial synergies. NBER Working Paper.

Kline, P., and E. Moretti. 2014b. People, places, and public policy: Some simple welfare economics of local economic development programs. *Annual Review of Economics* 6:629–62.

———. 2014a. People, places, and public policy: Some simple welfare economics of local economic development programs. *Annual Review of Economics* 6:629–62.

Leduc, S., and D. Wilson. 2012. Roads to prosperity or bridges to nowhere? theory and evidence on the effects of transportation infrastructure. In M. Eichenbaum, E. Hurst, and V. Ramey, eds., *NBER Macroeconomics Annual 2012, Volume 27*, 89–142. University of Chicago Press.

Lee, S. J., and M. C. Schmalz. 2026. Data centers and electricity bills: Pecuniary externality? Working Paper.

Liu, Y., J. Liu, and T. Wang. 2025. Large language model disagreement. Working Paper.

McFadden, D. 1978. Modeling the choice of residential location. In A. Karlqvist, L. Lundqvist, F. Snickars, and J. W. Weibull, eds., *Spatial Interaction Theory and Planning Models*, 75–96. Amsterdam: North-Holland.

Moretti, E., and D. J. Wilson. 2014. State incentives for innovation, star scientists and jobs: Evidence from biotech. *Journal of Urban Economics* 79:20–38.

———. 2017. The effect of state taxes on the geographical location of top earners: Evidence from star scientists. *American Economic Review* 107:1858–903.

Sheng, J., Z. Sun, B. Yang, and A. L. Zhang. (forthcoming). Generative ai and asset management. *Review of Financial Studies* .

Suárez Serrato, J. C., and O. Zidar. 2016. Who benefits from state corporate tax cuts? a local labor markets approach with heterogeneous firms. *American Economic Review* 106:2582–624.

Suarez Serrato, J. C., and O. Zidar. 2016. Who benefits from state corporate tax cuts? A local labor markets approach with heterogeneous firms. *American Economic Review* 106:2582–624.

Sun, L., and S. Abraham. 2021. Estimating dynamic treatment effects in event studies with heterogeneous treatment effects. *Journal of Econometrics* 225:175–99.

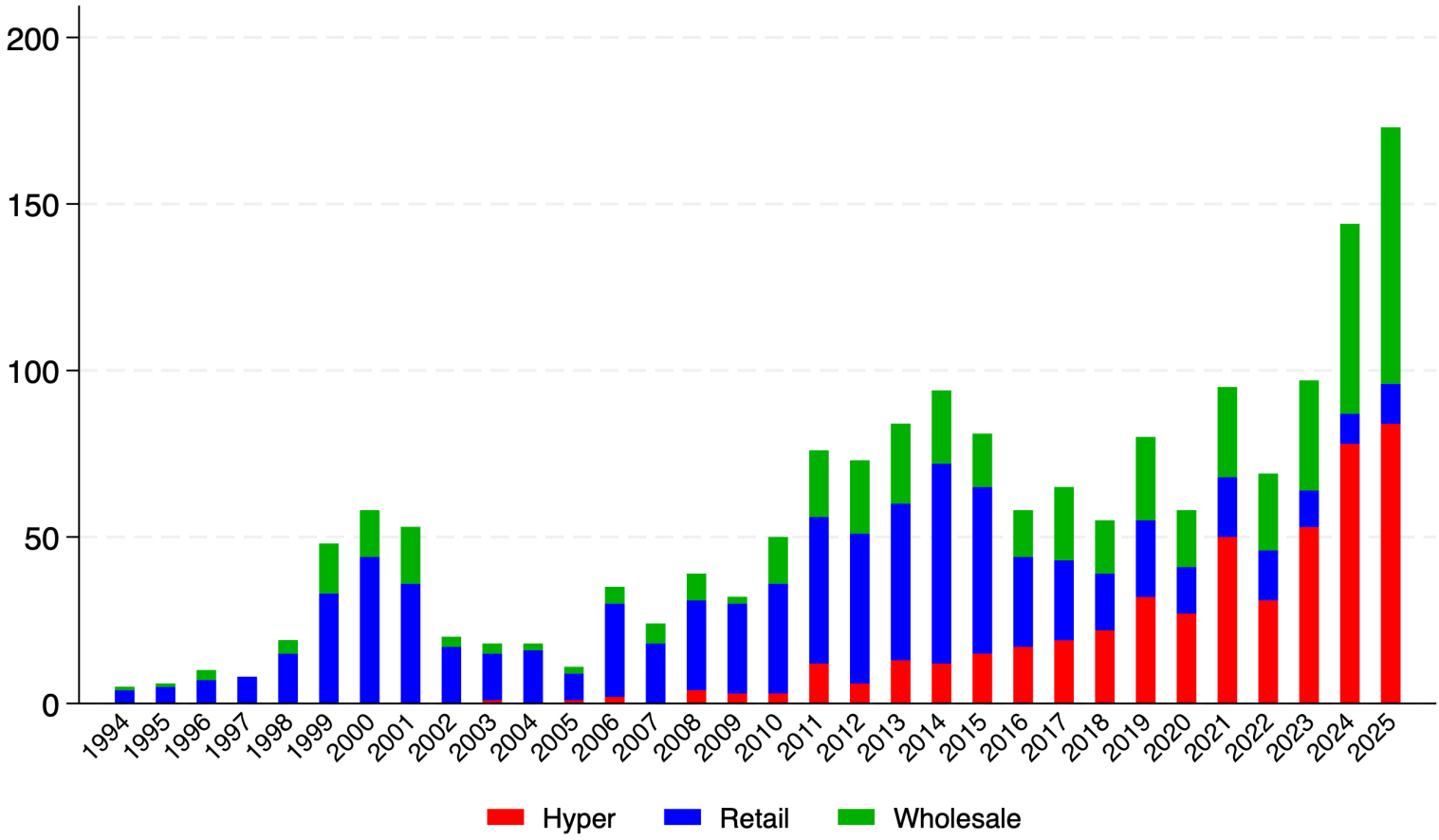


Figure 1: This Figure displays the number of new data centers built in the U.S. in each year from 1994 to 2025. Blue, green, and red stacked bars refer to retail, wholesale, and hyperscale data centers, respectively.

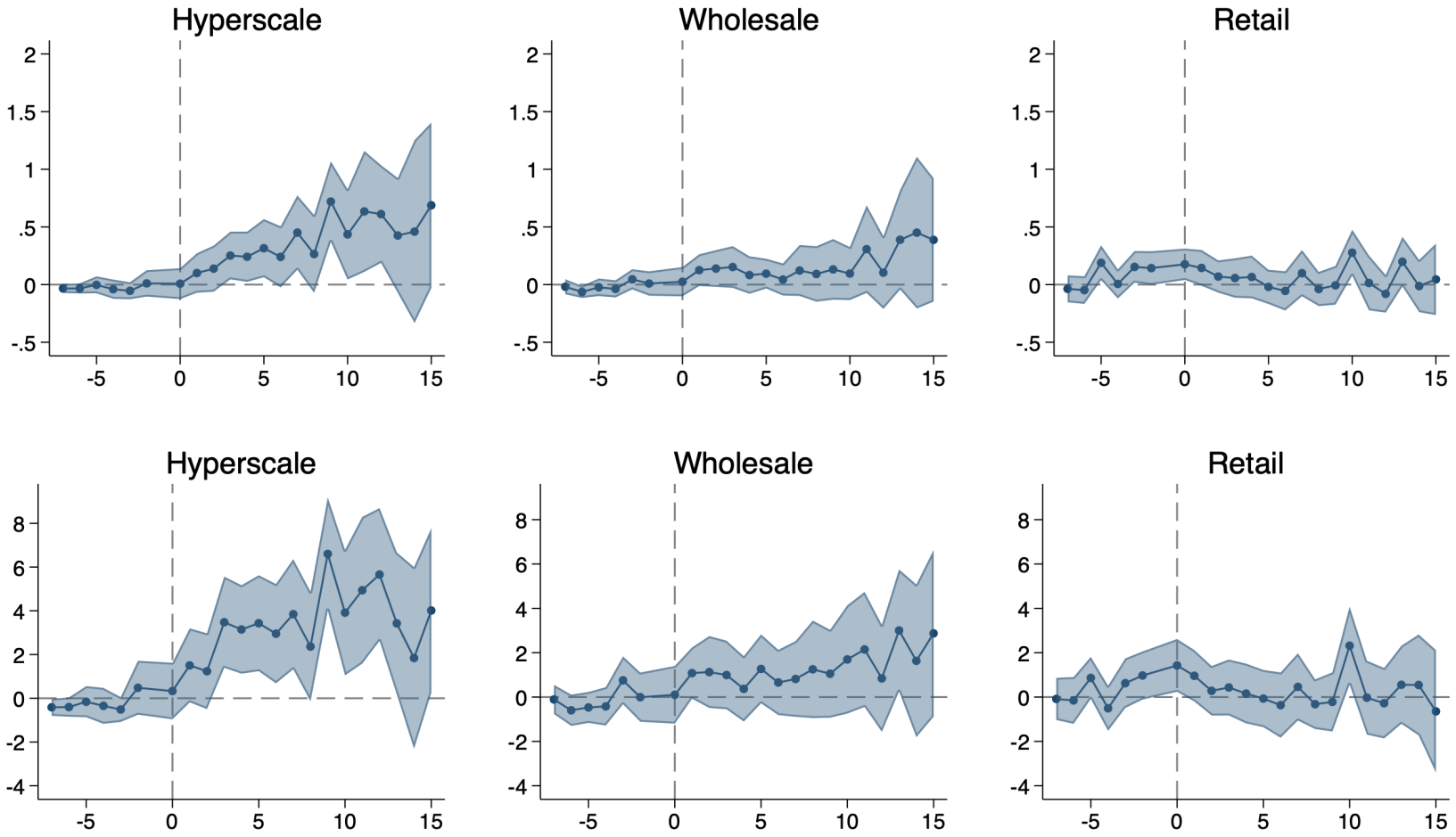


Figure 2: This Figure displays estimates of dynamic treatment effects coefficients based on the regression approach in Sun and Abraham (2021). On the x-axis, the value of zero corresponds to the year of implementation of a tax incentive policy, negative numbers correspond to years before implementation, and positive numbers to years after the implementation. For the panels in the top row, the dependent variable is the log number of data centers completed in each year. For the panels in the bottom row, the dependent variable the log total UPS power capacity from the completed data centers. We display results separately for hyperscale, wholesale, and retail data centers. The sample covers the years from 1994 to 2025.

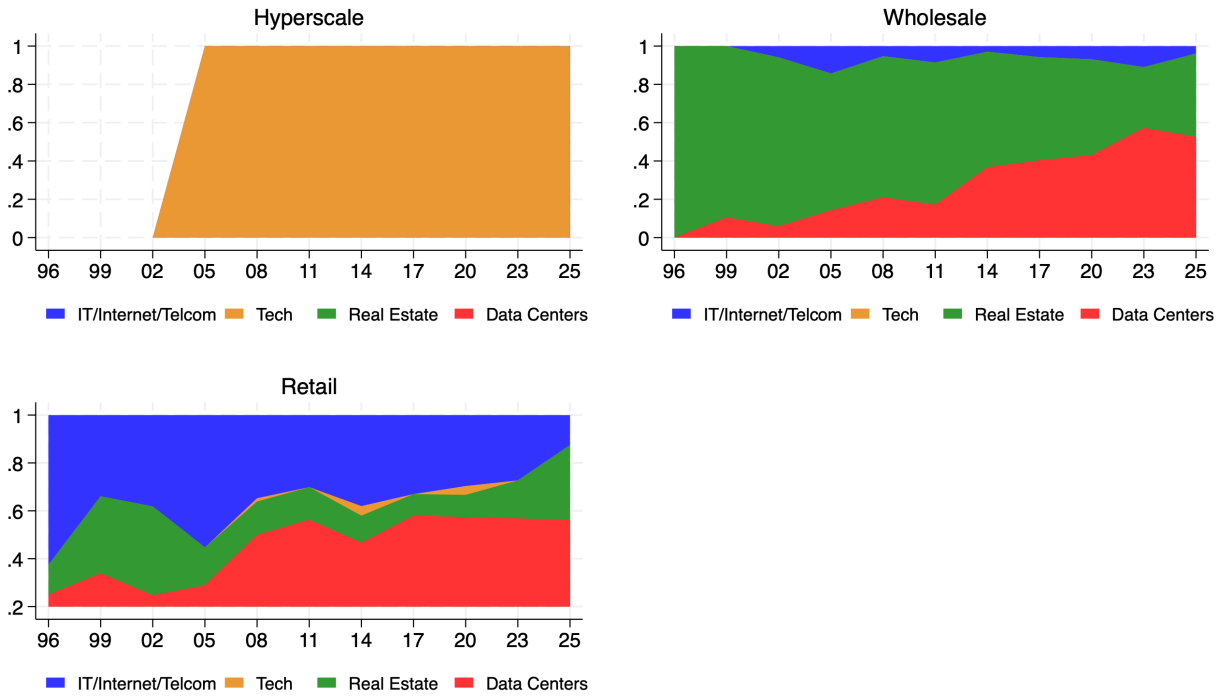
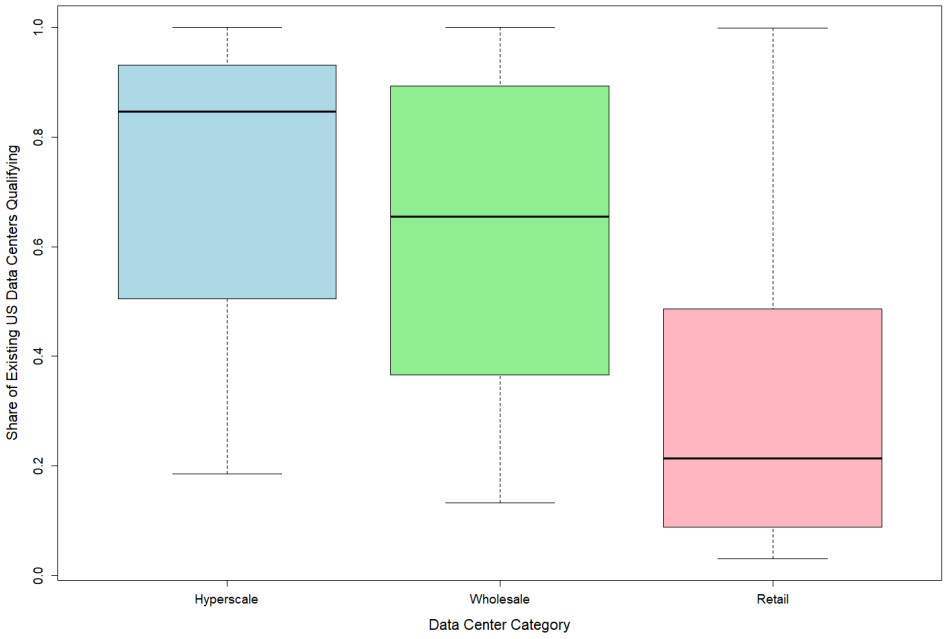
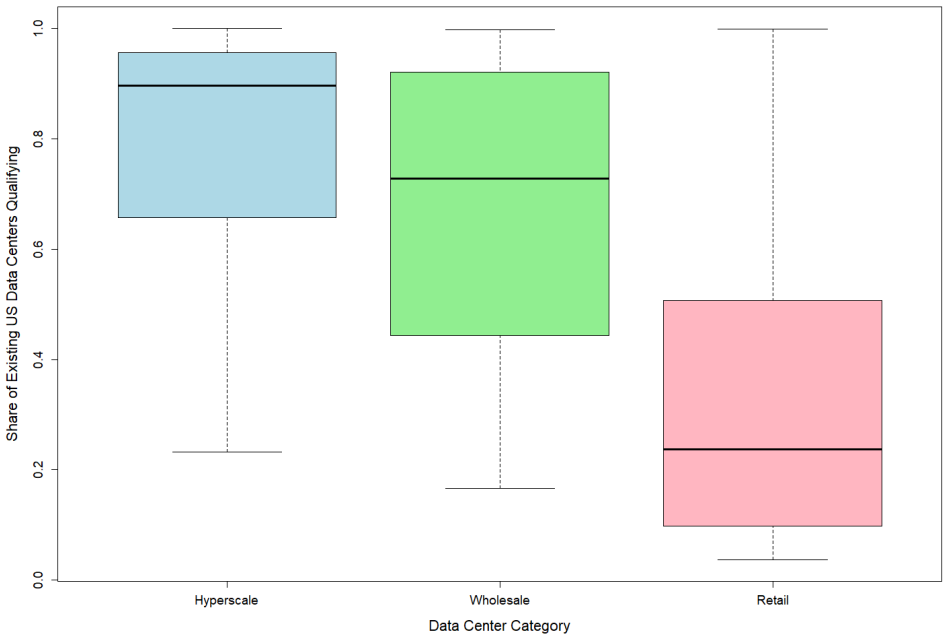


Figure 3: This Figure displays ownership composition over time for data centers completed in each year from 1996 to 2025. The three panels correspond to hyperscale (upper-left plot), wholesale (upper-right), and retail (lower-right) data centers. *IT/Internet/Telcom* stands for companies that specialize in IT services and network providers. *Tech* stands for large tech companies (e.g. Microsoft or Amazon). *Real Estate* stands for real estate investors (REITs, private equity funds, real estate operating companies). *Data Centers* stands for LLCs owning a limited number of (frequently a single) data center. It is likely that companies in this last group are owned by larger real estate investors.



(a) Existing Data Center Stock in 2020



(b) Existing Data Center Stock in 2025

Figure 4: This figure provides a graphical representation of the extent to which minimum investment size requirements for state-level tax incentives are binding across data center categories (hyperscale, wholesale, and retail). For each state's requirement, we compute the share of the total U.S. data center stock in 2020 (panel a) and 2025 (panel b) that would qualify under that threshold. The distribution of these shares across states and for each data center category is displayed using box plots.

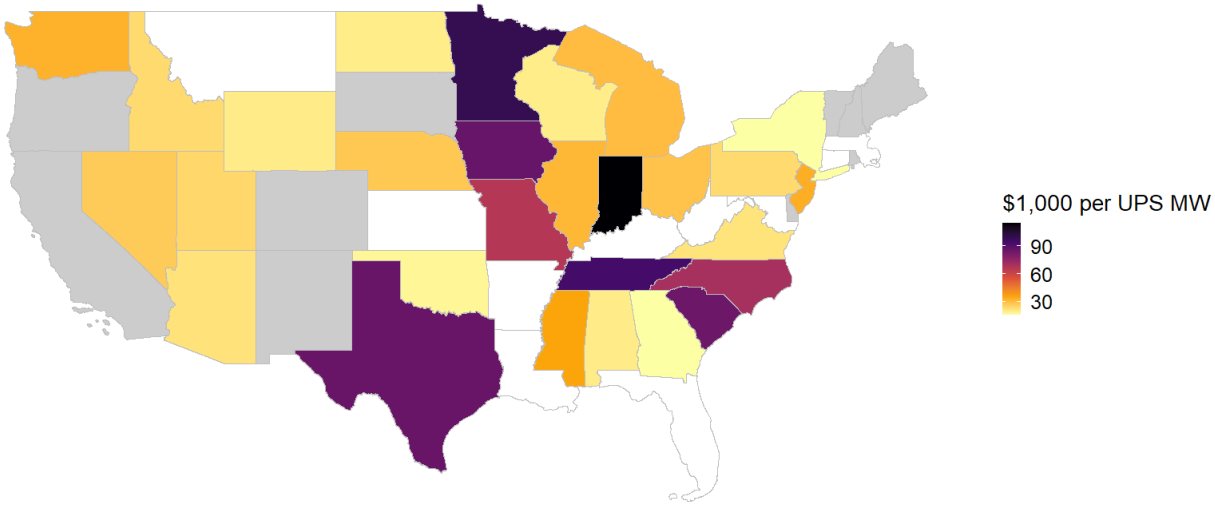


Figure 5: This Figure reports the average tax savings per megawatt of UPS power capacity, across states, and for data centers that qualify for state-level tax incentives and are completed and operational as of Q2 2025. The estimates reflect savings from sales and use tax incentives in 2025. States in white have sale and use tax incentive provisions in place but no qualifying data centers in our dataset, while states in grey do not have incentives.

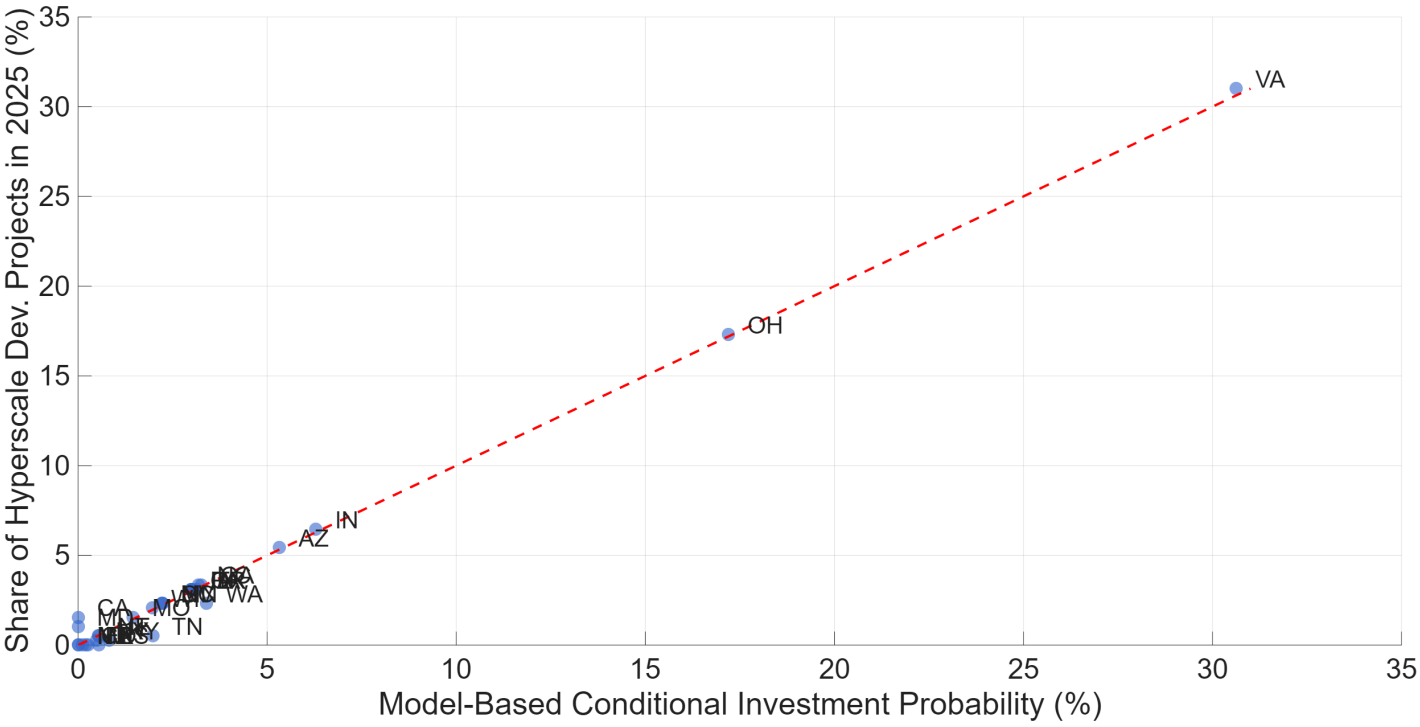
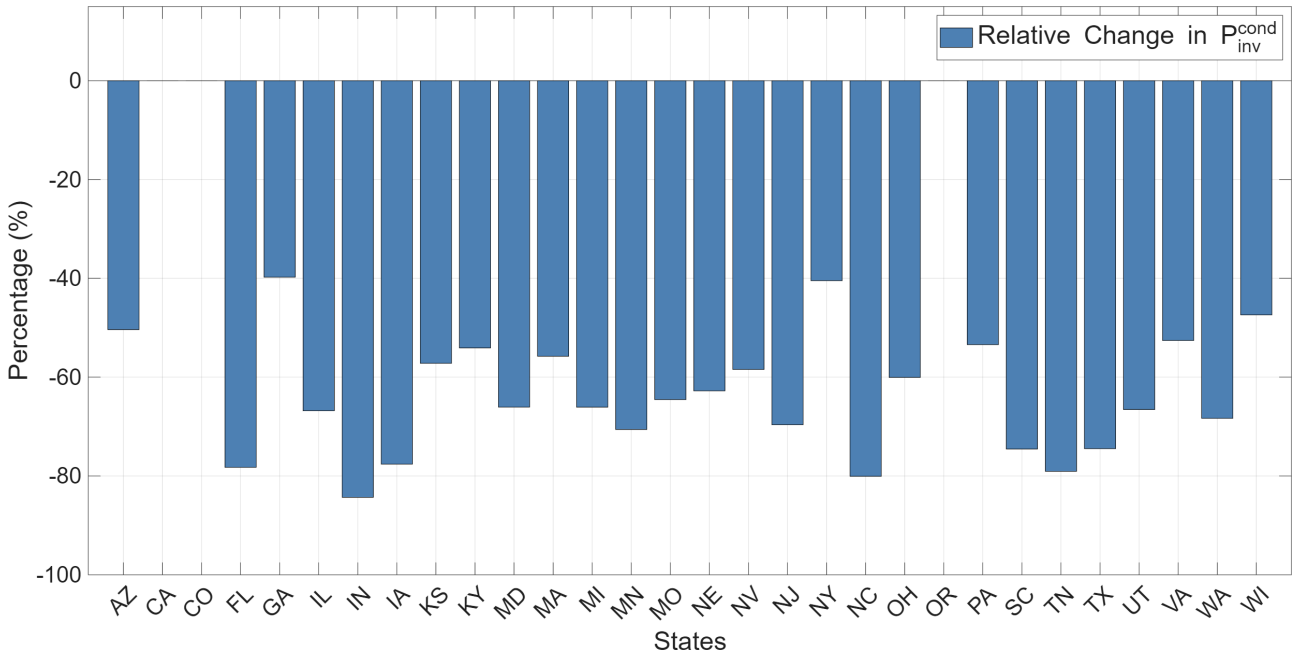
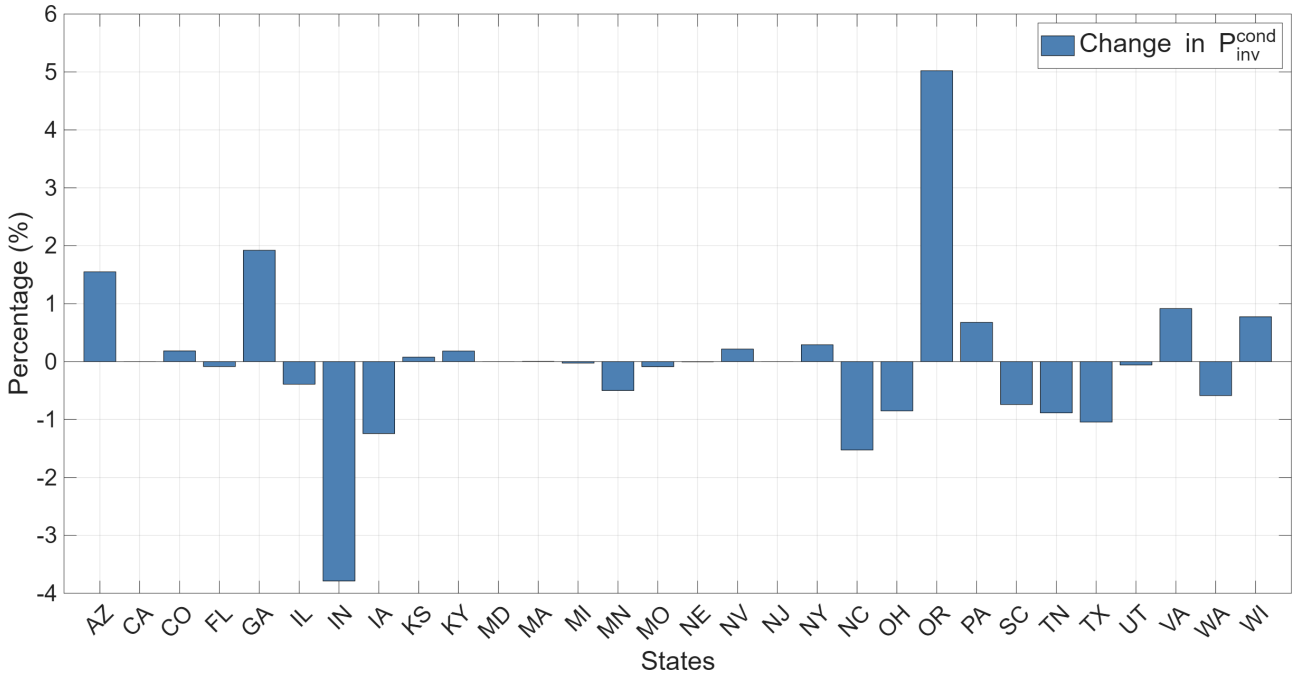


Figure 6: This Figure displays the fit of the model to the distribution of new hyperscale data center projects across U.S. states. Each point corresponds to one of the 30 states included in the model calibration. The x-axis displays values of the model-based conditional investment probability (P_{inv}^{cond}, see equation 7), while the y-axis displays value of the share of new hyperscale data center projects (share of under construction and planned projects) in 2025. The dashed line is the 45-degree line.



(a) State-Level Counterfactual: Relative Change in Conditional Investment Probabilities ($P_{inv}^{condNI}/P_{inv}^{cond} - 1$)



(b) U.S.-Level Counterfactual: Change in Conditional Investment Probabilities ($P_{inv}^{condNI} - P_{inv}^{cond}$)

Figure 7: The Figure displays the redistribution of new hyperscale data center projects in two counterfactual exercises. The results focus on changes in conditional investment probabilities in a counterfactual with no tax incentives (P_{inv}^{condNI}), compared to the baseline (P_{inv}^{cond}). Panel (a) displays the relative change in conditional investment probabilities for each state, in an exercise in which tax incentives are removed only for that state. Panel (b) shows changes in conditional investment probabilities across states in a counterfactual in which incentives are removed for all states.

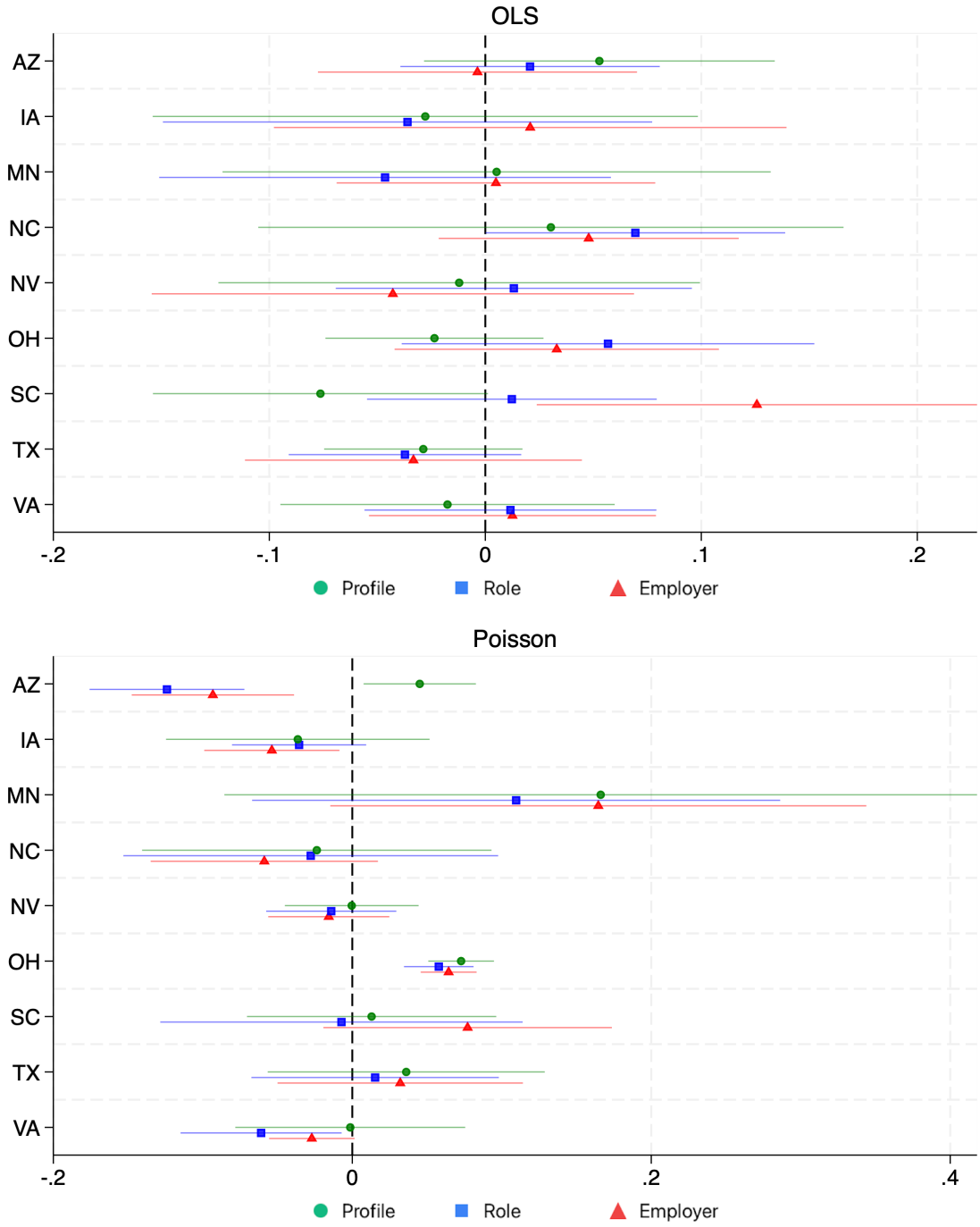


Figure 8: This Figure displays coefficient estimates and 95% confidence intervals from the following regressions

$$log(N_{TechJobs,c,t+1} + 1) = \beta z_{DCenter,c,t} + a_c + a_t + e_{c,t}$$
$$N_{TechJobs,c,t+1} = \exp(\psi z_{DCenter,c,t} + p_c + p_t + u_{c,t})$$

where $N_{TechJobs,c,t+1}$ is the number of new tech jobs in census place c and year $t+1$; $z_{DCenter,c,t}$ is a dummy equal to one if at least one new data center is completed within 10 miles of census place or city c in year t; a_c and a_t (p_c and p_t) are census place and year fixed effects. Green circles represent estimates from specifications in which the number of new tech jobs is measured using profile description, blue squares represent estimates in which it is measured using standardized roles, and red circles represent estimates in which it is measured using employer characteristics. Estimates are based on a matched sample over the years from 2013 to 2025, for each state. In the matching approach, each exposed unit (i.e., a city or census place with at least one occurrence of $z_{DCenter,c,t} = 1$) is matched with one non-exposed unit (i.e., a city or census place with $z_{DCenter,c,t} = 0$ over the entire sample).

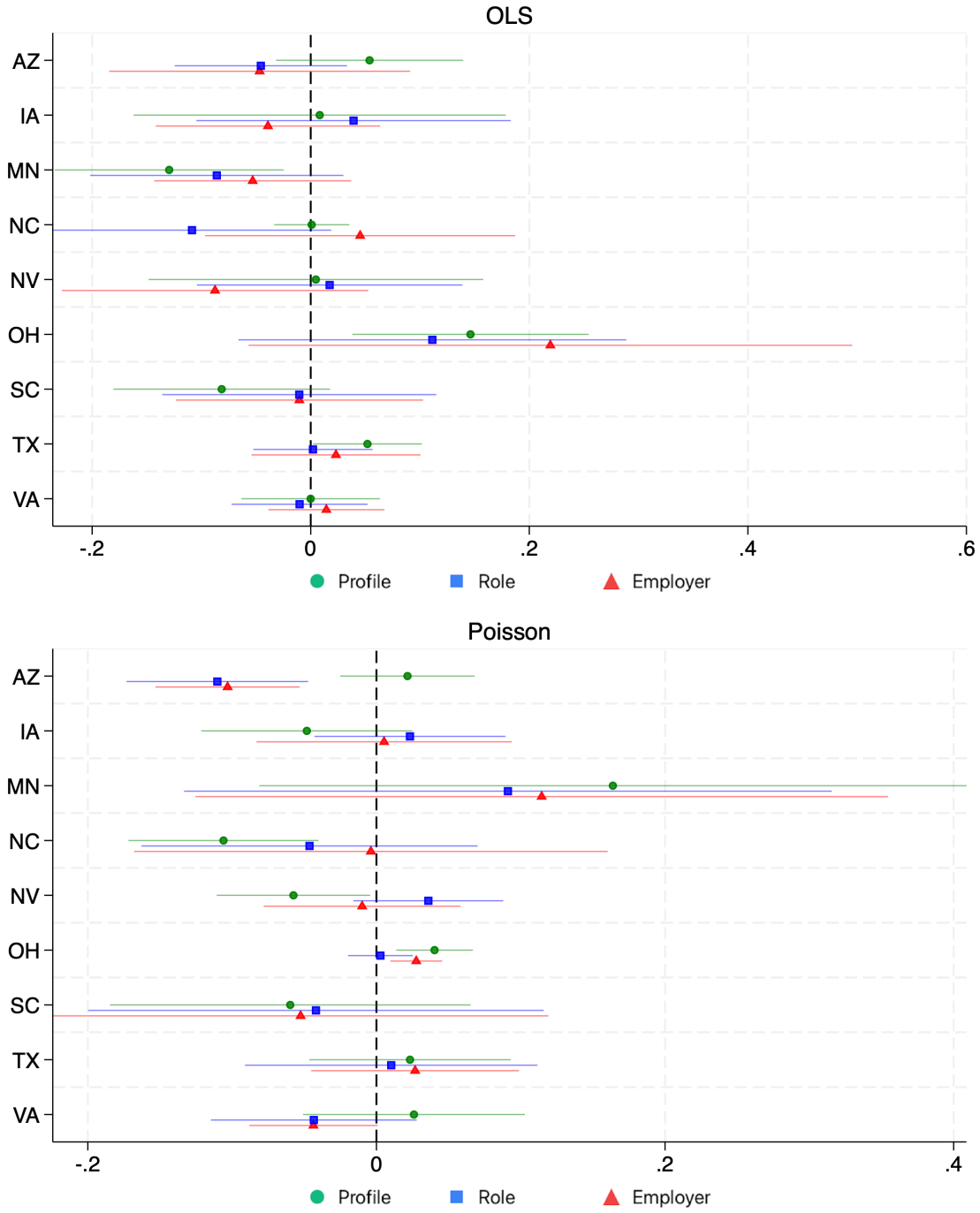


Figure 9: This Figure displays coefficient estimates and 95% confidence intervals from the following regressions

$$log(N_{TechJobs,c,t+3}+1) = \beta z_{DCenter,c,t} + a_c + a_t + e_{c,t}$$
$$N_{TechJobs,c,t+3} = \exp(\psi z_{DCenter,c,t} + p_c + p_t + u_{c,t})$$

where $N_{TechJobs,c,t+3}$ is the number of new tech jobs in census place or city c and year $t+3$; $z_{DCenter,c,t}$ is a dummy equal to one if at least one new data center is completed within 10 miles of census place c in year t; a_c and a_t (p_c and p_t) are census place and year fixed effects. Green circles represent estimates from specifications in which the number of new tech jobs is measured using profile description, blue squares represent estimates in which it is measured using standardized roles, and red circles represent estimates in which it is measured using employer characteristics. Estimates are based on a matched sample over the years from 2013 to 2025, for each state. In the matching approach, each exposed unit (i.e., a city or census place with at least one occurrence of $z_{DCenter,c,t} = 1$) is matched with one non-exposed unit (i.e., a city or census place with $z_{DCenter,c,t} = 0$ over the entire sample).

Table 1: Summary Statistics

	A: Hyperscale					
	N	Mean	SD	p25	p50	p75
Power (MW)	485	33.23	26.73	17.20	30	45
Size (1,000 SF)	485	168.60	125.81	98.83	157.69	218.67
Num. Racks (100)	485	48.54	36.27	28.57	45.06	63.81
Single	485	0.95	0.21	1	1	1
Pop. Density	483	646	854	68	256	1,000
Rural	485	0.17	0.37	0	0	0
Non Metro	485	0.27	0.45	0	0	1
	B: Wholesale					
	N	Mean	SD	p25	p50	p75
Power (MW)	516	18.04	22.84	4	11.72	24.59
Size (1,000 SF)	516	103.25	113.95	28.20	79.39	148.38
Num. Racks ('00)	516	29.67	32.47	8.21	22.86	42.86
Single	516	0.17	0.37	0	0	0
Pop. Density	516	3,602	9,925	987	1,652	2,794
Rural	516	0.04	0.20	0	0	0
Non Metro	516	0.05	0.21	0	0	0
	C: Retail					
	N	Mean	SD	p25	p50	p75
Power (MW)	756	3.90	6.94	0.35	1.50	4.33
Size (1,000 SF)	756	27.24	41.84	3.18	10	32.19
Num. Racks (100)	756	7.85	12.32	0.90	2.89	9.36
Single	756	0.04	0.21	0	0	0
Pop. Density	756	4,798	8,542	1,152	2,180	4,351
Rural	756	0.01	0.07	0	0	0
Non Metro	756	0.01	0.12	0	0	0

This Table displays cross-sectional summary statistics across the main three types of data centers in our dataset: hyperscale (panel A), wholesale (panel B), and retail (panel C). We report the mean, standard deviation, as well as the $25th$, $50th$, and $75th$ percentiles. *Power*, is the total uninterruptible power supply (UPS) capacity in megawatts; *Size*, is the operational space in thousands of square feet; *Racks*, is the number of racks in hundreds; *Is Single*, is a dummy variable equal to one if the data center is single tenant; *Pop Density*, the population density of the census track where the data center is located; *Rural*, is a dummy variable equal to one if the data center is located in a rural census track; *Non Metro*, is a dummy variable equal to one if the data center is located outside a metropolitan area.

Table 2: Tax Incentives Legislation

State	Year	Type	Coverage	Minimum Requirements	Duration (years)	Other
Alabama	2012	Sales, Use & Property	Equipment, infrastructure	Tiers: between 200M & 400M	Between 10 and 30	20 jobs; 40k avg wage
Arizona	2013	Sales & Use	Equipment	Tiers: between 50M & 25M (on population)	Between 10 and 20	
Arkansas	2023	Sales & Use	Equipment, electricity	500M		1M in aggregate wage
Connecticut	2021	Sales, Use & Property	Equipment, electricty	Tiers: between 50M and 400M (on location)	Between 20 and 30	
Florida	2017	Sales, Use & Property	Equipment, electricty	150M		15 Megawatt
Georgia	2018	Sales & Use	Equipment	250M		
Idaho	2020	Sales, Use & Property	Equipment	250M		30 jobs
Illinois	2019	Sales & Use	Equipment, infrastructure	250M	20	20 jobs; 120% of county median wage
Indiana	2019	Sales & Use	Equipment, electricity	Tiers: between 25M and 150M (on population)	25	
Iowa	2009	Sales & Use	Equipment, electricity	200M		
Kansas	2025	Sales & Use	Equipment, infrastructure	250M	20	20 jobs
Kentucky	2024	Sales & Use	Equipment	Tiers: between 25M and 450M (on population)	Between 25 and 50	
Louisiana	2024	Sales & Use	Equipment, infrastructure	200M	20	50 jobs
Maryland	2020	Sales & Use & Property	Equipment, infrastructure	Tiers: between 2M and 250M (on location)	Between 10 and 20	5 jobs; 150% of county median wage
Massachusetts	2025	Sales & Use	Equipment, electricity	50M	20	100 jobs
Michigan	2016	Sales & Use	Equipment, infrastcture			
Minnesota	2011	Sales & Use	Equipment, infastructure, electricity	Tiers: between 30M and 50M	20	Between 25k sqft and 30k sqft
Mississippi	2010	Sales & Use	Equipment, infrastcture	50M		50 jobs; 150% of state median wage
Missouri	2015	Sales & Use	Equipment, electricity	Tiers: between 5M and 25M	Between 10 and 15	Between 5 and 10 jobs
Montana	2018	Sales, Use & Property	Equipment, infrastructure	150M and 300k sqft	15	
Nebraska	2012	Sales, Use & Property	Equipment, infrastructure	Tiers: between 1M and 100M	Varies by tier	30 employees (Tier 1)
Nevada	2015	Sales, Use & Property	Equipment	Tiers: between 25M and 100M	Between 10 and 20	10 jobs at average wages
New Jersey	2024	Sales & Use	Equipment, infrastructure	100M	10	100 jobs; 120% of county median wage
New York	2000	Sales & Use	Equipment	Specific for webhosting		
North Carolina	2010	Sales & Use	Equipment, electricity	Tiers: between 150M and 225M		
North Dakota	2015	Sales & Use	Equipment	15k sqft		
Ohio	2013	Sales, Use & Property	Equipment, infrastructure	100M		1.5M in aggregate wages
Oklahoma	2012	Sales & Use	Equipment			Majority of revenue from out-of-state
Pennsylvania	2022	Sales & Use	Equipment	Tiers: Between 75M and 100M (on population)	25	25 jobs and 1M total compensation
South Carolina	2012	Sales, Use & Property	Equipment, electricity	50M		25 jobs
Tennessee	2016	Sales & Use	Equipment, electricity	250M		15 jobs; 150% of average wage
Texas	2013	Sales & Use	Equipment, electricity	Tiers: between 200M and 250M	Between 10 and 15	20 jobs; above-average wages and 100k sqft
Utah	2020	Sales & Use	Equipment, infrastructure	case by case		
Virginia	2010	Sales & Use	Equipment, infrastructure	Tiers: between 70M and 150M (on location)	25	Between 10 and 50 jobs; 150% of avg wage
Washington	2010	Sales & Use	Equipment, infrastructure	100k sqft		35jobs; 150% of per capital personal income
West Virginia	2025	Sales & Use	Equipment	50M		50 jobs
Wisconsin	2023	Sales & Use	Equipment	Tiers: between 50M and 150M (on population)		
Wyoming	2010	Sales & Use	Equipment, infrastructure	5M		

This Table summarizes the main features of the tax incentives provisions for new data center development available in each state. The column *State* lists the state enacting the provision; *Year* indicates the year the provision first became effective; *Type* describes the types of incentives offered; *Coverage* specifies the items covered by sales and use tax exemptions; *Minimum Requirements* shows the minimum investment (in millions of dollars) or facility size (in square feet) required to qualify for the incentives; *Duration* shows the maximum number of years after facility completion for which tax benefits are available; *Other* includes additional details, such as requirements on new job creation.

Table 3: Tax Incentives and Data Center Development

A: Hyperscale

	$\mathbb{1}\{Built\}$		$ln(1+N)$		N		$ln(1+Power)$	
D	0.26***	0.26***	0.26***	0.26***	1.54***	1.52***	2.75***	2.76***
	(4.64)	(4.71)	(3.12)	(3.22)	(4.24)	(5.06)	(4.38)	(4.45)
Controls	NO	YES	NO	YES	NO	YES	NO	YES
R^2-adj	0.396	0.400	0.423	0.426	-	-	0.397	0.400
N	1632	1581	1632	1581	525	480	1632	1581

B: Wholesale

	$\mathbb{1}\{Built\}$		$ln(1+N)$		N		$ln(1+Power)$	
D	0.09**	0.08**	0.11**	0.11*	0.33	0.29	0.92**	0.86**
	(2.25)	(2.19)	(2.10)	(2.00)	(0.95)	(1.13)	(2.17)	(2.10)
Controls	NO	YES	NO	YES	NO	YES	NO	YES
R^2-adj	0.393	0.390	0.469	0.468	-	-	0.415	0.415
N	1632	1581	1632	1581	1023	990	1632	1581

C: Retail

	$\mathbb{1}\{Built\}$		$ln(1+N)$		N		$ln(1+Power)$	
D	0.02	0.03	0.03	0.04	0.14	0.15	0.35	0.41
	(0.68)	(0.97)	(0.84)	(0.95)	(0.95)	(1.05)	(1.29)	(1.52)
Controls	NO	YES	NO	YES	NO	YES	NO	YES
R^2-adj	0.373	0.373	0.413	0.414	-	-	0.371	0.370
N	1632	1581	1632	1581	1408	1364	1632	1581

This Table displays coefficient estimates (and associated t-stats) for the following regression equation:

$$y_{s,t} = \delta D_{s,t} + BX_{s,t} + a_s + a_t + e_{s,t}$$

where $y_{s,y}$ denotes a dummy equal to one if at least one new data center is built in state s and year t, the log of the number of new data centers built in state s and year t, the number of new data centers built in state s and year t, the log of total UPS (in kW) of new data centers built in state s and year t; $D_{s,t}$ is a dummy variable equal to one if tax incentives are available in state s and year t; $X_{s,t}$ is a vector of state-level time-varying controls; a_s and a_t are state and year fixed effects. When the dependent variable is the number of built data centers, estimates are based on Poisson regressions. Estimates reported in panel A (B) (C) are for the sample of hyperscale (wholesale) (retail) data centers. T-stats are based on standard errors double-clustered by state and year. Coefficients marked with ***, **, and * are significant at the 1%, 5%, and 10% level.

Table 4: Tax Incentives and Data Center Development: Alternative Estimators

A: Stacked Regression

	ln(1+N)			ln(1+Power)		
	Hyper	*Wholesale*	*Retail*	*Hyper*	*Wholesale*	*Retail*
D	0.24***	0.11*	0.03	2.55***	0.96**	0.31
	(2.95)	(1.95)	(0.99)	(4.29)	(2.20)	(1.17)
R^2-adj	0.443	0.540	0.559	0.432	0.536	0.524
N	8288	8288	8288	8288	8288	8288

B: Sun & Abraham (2021)

	ln(1+N)			ln(1+Power)		
D	0.27***	0.12**	0.04	2.81***	0.93**	0.34
	(3.45)	(2.19)	(1.07)	(4.93)	(2.25)	(1.67)
R^2-adj	0.443	0.540	0.559	0.432	0.536	0.524
N	1632	1632	1632	1632	1632	1632

C: Callaway & Sant'Anna (2021)

	ln(1+N)			ln(1+Power)		
D	0.28**	0.01	-0.22**	2.62***	-0.32	-1.28**
	(2.38)	(0.02)	(-2.51)	(2.89)	(-0.46)	(-2.13)
R^2-adj	0.443	0.540	0.559	0.432	0.536	0.524
N	1632	1632	1632	1632	1632	1632

This Table displays coefficient estimates (and associated t-stats) for the following regression equation:

$$y_{s,t} = \delta D_{s,t} + a_s + a_t + u_{s,t}$$

where $y_{s,t}$ denotes the number of new data centers built in state s and year t, or the log of total UPS (in kW) of new data centers built in state s and year t; $D_{s,t}$ is a dummy variable equal to one if tax incentives are available in state s and year t; a_s and a_t are state and year fixed effects. Estimates in columns 1 and 4 (2 and 5) (3 and 6) are for hyperscale (wholesale) (retail) data centers. Results in Panels A, B, and C are based on the estimators in Cengiz et al. (2019), Sun and Abraham (2021) and Callaway and Sant'Anna (2021), respectively. T-stats are based on standard errors double-clustered by state and year. Coefficients marked with ***, **, and * are significant at the 1%, 5%, and 10% level.

Table 5: Tax Incentives and Data Center Development: Incentive Scope Scores

A: Hyperscale

	$\mathbb{1}\{Built\}$		$ln(1+N)$		N		$ln(1+Power)$	
IncentivesScope	0.08***	0.08***	0.09***	0.09***	0.40***	0.40***	0.87***	0.85***
	(5.07)	(5.01)	(3.27)	(3.28)	(3.97)	(4.73)	(4.74)	(4.70)
R^2-adj	0.412	0.415	0.442	0.444			0.415	0.416
Obs	1632	1581	1632	1581	525	480	1632	1581

B: Wholesale

	$\mathbb{1}\{Built\}$		$ln(1+N)$		N		$ln(1+Power)$	
IncentivesScope	0.03**	0.03**	0.04**	0.04**	0.09	0.08	0.31**	0.30**
	(2.49)	(2.48)	(2.18)	(2.09)	(1.11)	(1.31)	(2.40)	(2.36)
R^2-adj	0.395	0.392	0.474	0.473			0.418	0.419
Obs	1632	1581	1632	1581	1023	990	1632	1581

C: Retail

	$\mathbb{1}\{Built\}$		$ln(1+N)$		N		$ln(1+Power)$	
IncentivesScope	0.01	0.01	0.02	0.02	0.06	0.06	0.14*	0.15**
	(1.15)	(1.42)	(1.32)	(1.42)	(1.55)	(1.63)	(1.84)	(2.07)
R-Square	0.373	0.374	0.413	0.414			0.372	0.371
Obs	1632	1581	1632	1581	1408	1364	1632	1581

This Table displays coefficient estimates (and associated t-stats) for the following regression equation:

$$y^k_{s,t} = \gamma\big(D_{s,t} \times IncentivesScope_s\big) + BX_{s,t} + a_s + a_t + v_{s,t}$$

where $y_{s,t}$ denotes a dummy equal to one if at least one new data center is built in state s and year t, the log of the number of new data centers built in state s and year t, the number of new data centers built in state s and year t, the log of total UPS (in kW) of new data centers built in state s and year t; $IncentivesScope_s$ is a state-level index of incentive scope and significance (see Internet Appendix B for more details); $X_{s,t}$ is a vector of state-level time-varying controls; a_s and a_t are state and year fixed effects. When the dependent variable is the number of built data centers, estimates are based on Poisson regressions. Estimates reported in panel A (B) (C) are for hyperscale (wholesale) (retail) data centers. T-stats are based on standard errors double-clustered by state and year. Coefficients marked with ***, **, and * are significant at the 1%, 5%, and 10% level.

Table 6: Tax Incentives and Data Center Development: Facility Characteristics

	A: Facility Characteristics								
	ln(Power)	*Single User*	*Cloud*	*Power D.*	*PUE*	*Office*	*On Site*	*ln(Pop. Den.)*	*Low Den.*
D	0.61***	0.20***	0.16***	0.01	-0.02	-0.07*	-0.07	-0.29	0.05*
	(3.02)	(3.31)	(3.02)	(0.97)	(-1.02)	(-2.01)	(-0.94)	(-1.18)	(1.75)
Controls	YES	YES	YES	YES	YES	YES	YES	YES	YES
State FE	YES	YES	YES	YES	YES	YES	YES	YES	YES
Year FE	YES	YES	YES	YES	YES	YES	YES	YES	YES
Category FE	NO	NO	NO	NO	NO	NO	NO	NO	NO
R-Square	0.363	0.295	0.177	0.237	0.090	0.144	0.359	0.311	0.259
Obs	1465	1470	1470	1465	1470	1470	1470	1463	1470
	B: Facility Characteristics (Control for Data Center Category)								
	ln(Power)	*Single User.*	*Cloud*	*Power D.*	*PUE*	*Office*	*On Site*	*ln(Pop. Den.)*	*Low Den.*
D	0.21*	0.02	0.02	0.00	-0.01	-0.04	-0.01	0.05	0.01
	(1.83)	(0.46)	(0.29)	(0.21)	(-0.59)	(-1.29)	(-0.12)	(0.24)	(0.57)
Controls	YES	YES	YES	YES	YES	YES	YES	YES	YES
State FE	YES	YES	YES	YES	YES	YES	YES	YES	YES
Year FE	YES	YES	YES	YES	YES	YES	YES	YES	YES
Category FE	YES	YES	YES	YES	YES	YES	YES	YES	YES
R-Square	0.587	0.703	0.485	0.269	0.499	0.176	0.425	0.400	0.298
Obs	1465	1470	1470	1465	1470	1470	1470	1463	1470

This Table displays coefficient estimates (and associated t-stats) for the following regression equation:

$$q_{i,t} = \phi D_{s,t} + BX_{s,t} + a_s + a_t + a_k + v_{i,t}$$

where $q_{i,t}$ is a characteristic of data center i built in year t (either the log of UPS, a dummy for single-user data centers, a dummy for data centers offering cloud services, the power usage effectiveness, a dummy for data centers with offices or on-site maintenance staff, log population density in the tract where the data center is located, or a dummy equal to one for data centers located in low population density tracts); $D_{s,t}$ is a dummy equal to one if tax incentives are available in state s and year t; $X_{s,t}$ is a vector of state-level time-varying controls; a_s and a_t are state and year fixed effects. T-stats are based on standard errors double-clustered by state and year. Coefficients marked with ***, **, and * are significant at the 1%, 5%, and 10% level.

Internet Appendix:

Subsidizing the Cloud: U.S. State Incentives to Data Centers

Appendix A: Additional Figures and Tables

Appendix B: Index of State Regulations Scope and Significance

Appendix C: Assumptions for State Tax Incentives Valuation

A Additional Figures and Tables

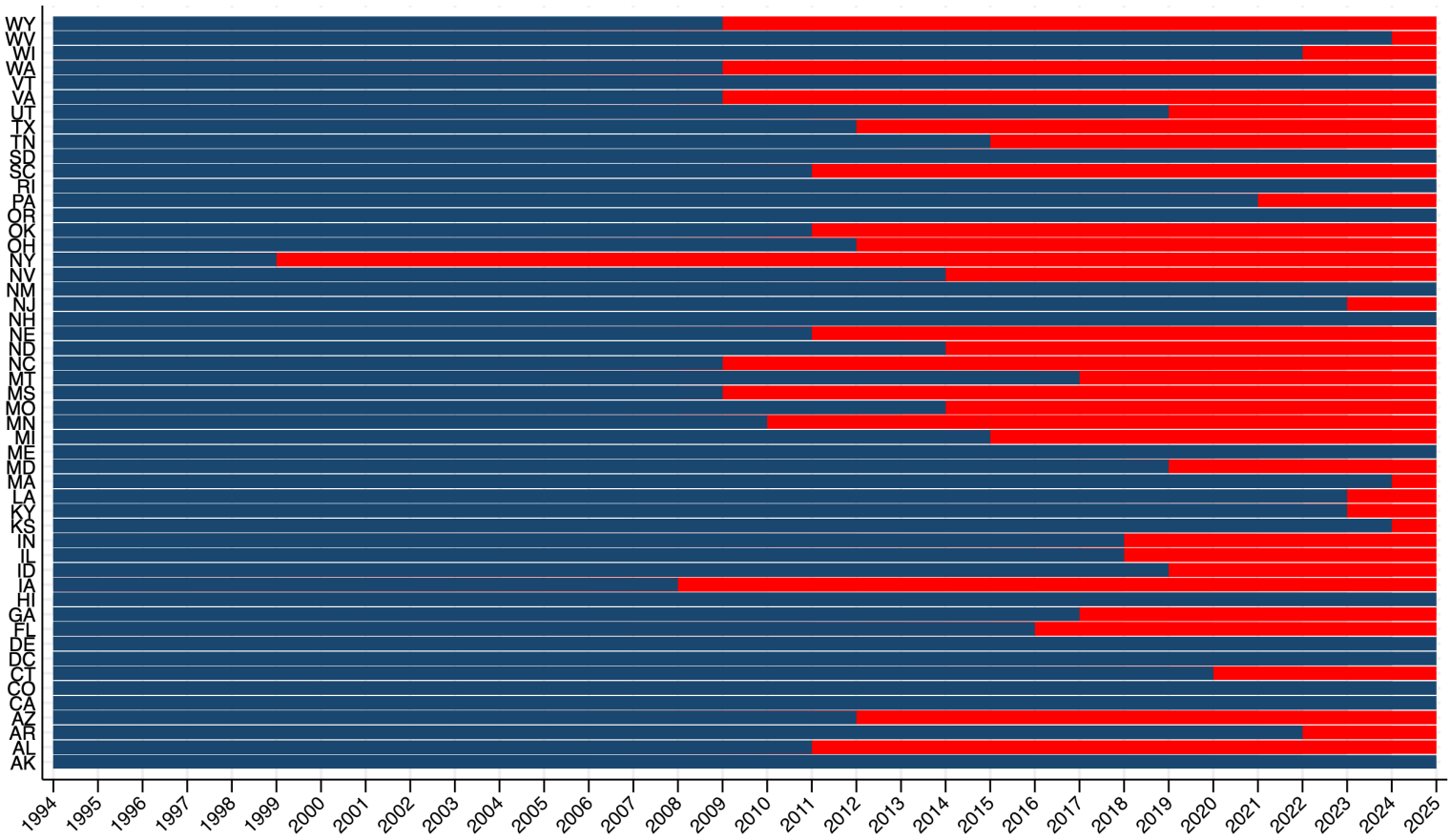
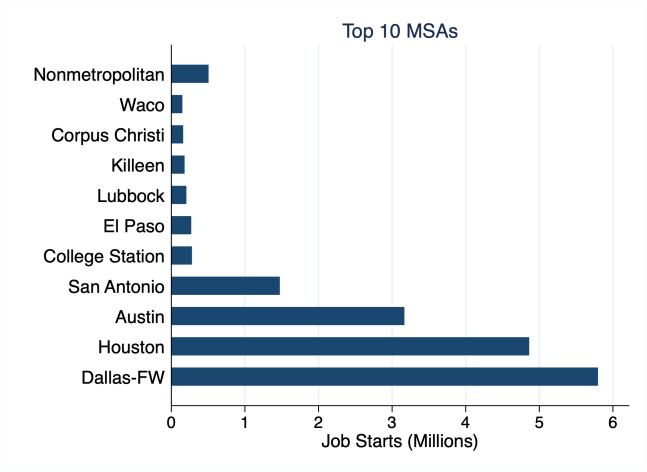
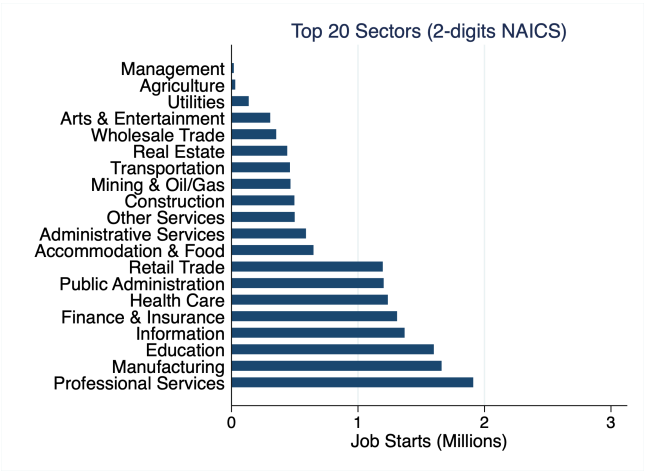


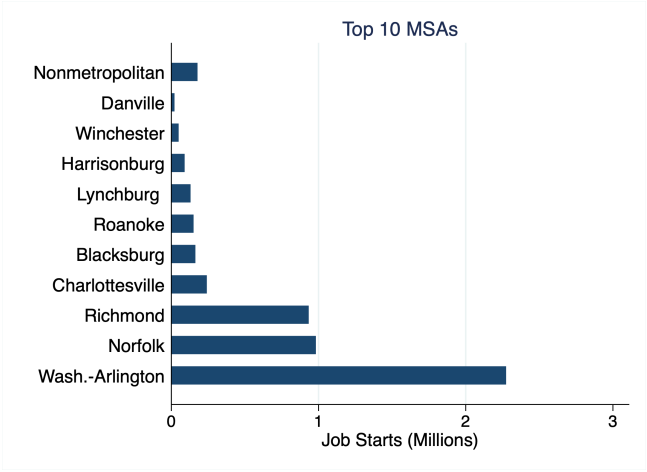
Figure A.1: This Figure provides a visual representation of the timing of staggered introduction of tax incentives across the U.S. states. The x-axis displays years from 1994 to 2025, while the y-axis lists all U.S. states. The horizontal bars are blue before the implementation of local tax incentives, and red after. States whose bars remain blue throughout the entire period never implemented incentives.



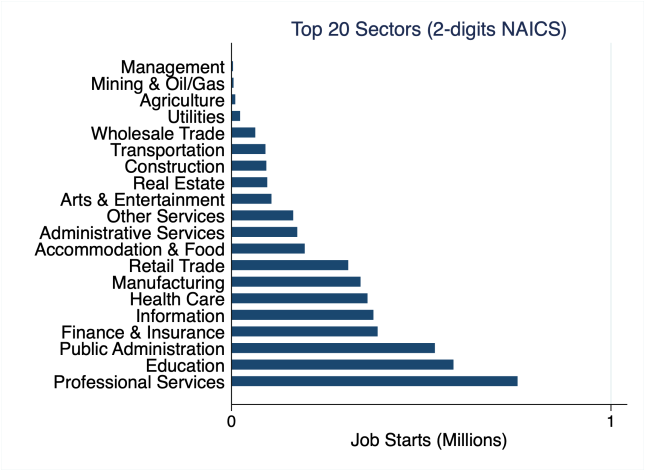
(a) Texas



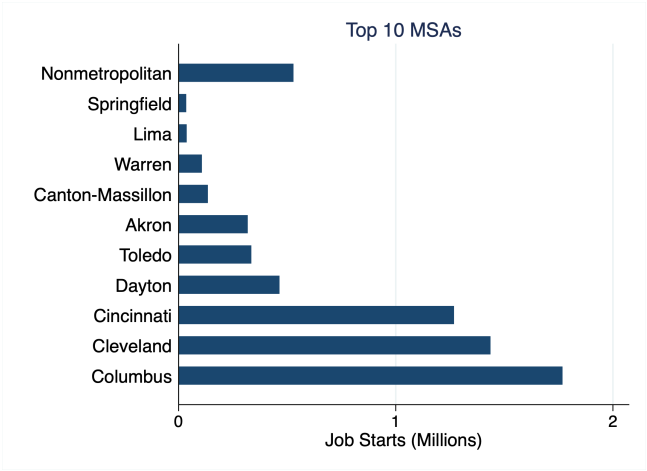
(b) Texas



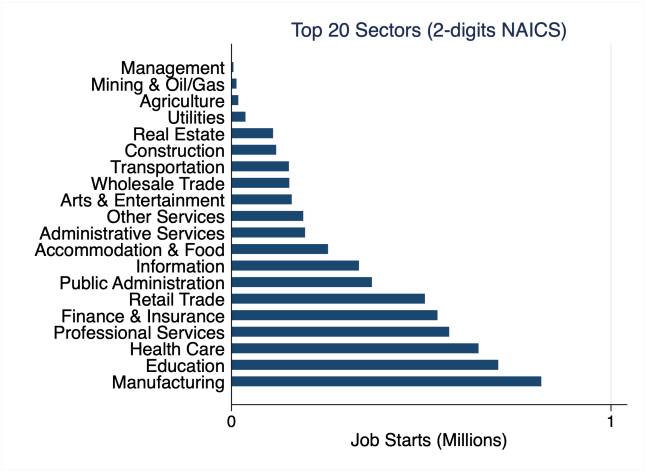
(c) Virginia



(d) Virginia



(e) Ohio



(f) Ohio

Figure A.2: This figure presents summary statistics from the RevelioLab dataset for Texas, Virginia, and Ohio. The panels on the left display total job starts from 2013 to 2025 in the top ten MSAs of each state, while the panels on the right report job starts for the top twenty industries.

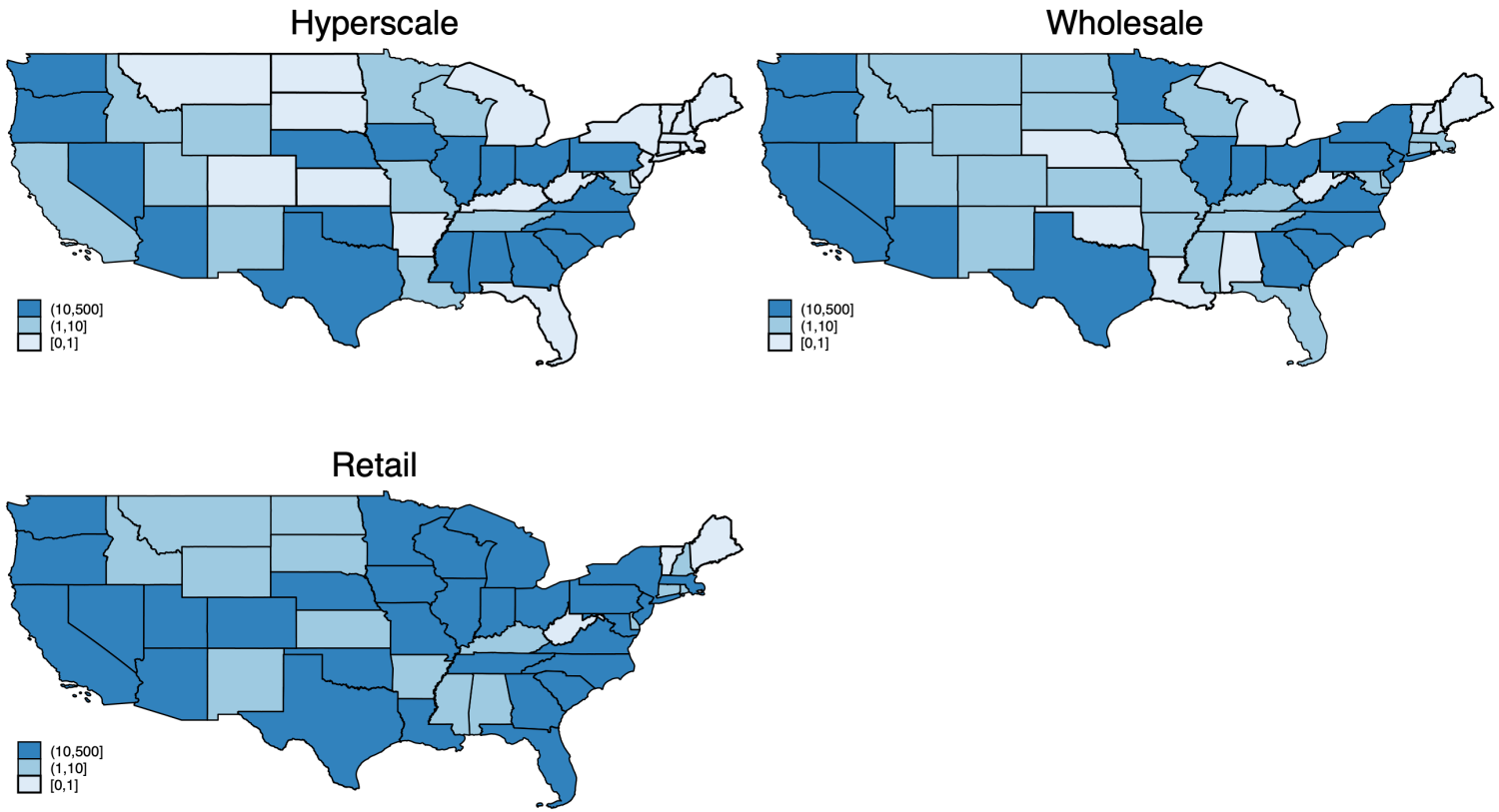


Figure A.3: This Figure displays a heatmap showing the number of operational, under construction, or planned data centers across U.S. states as of Q2 2025, and separately for hyperscale (top-left plot), wholesale (top-right plot) and retail (bottom-left plot) data centers.

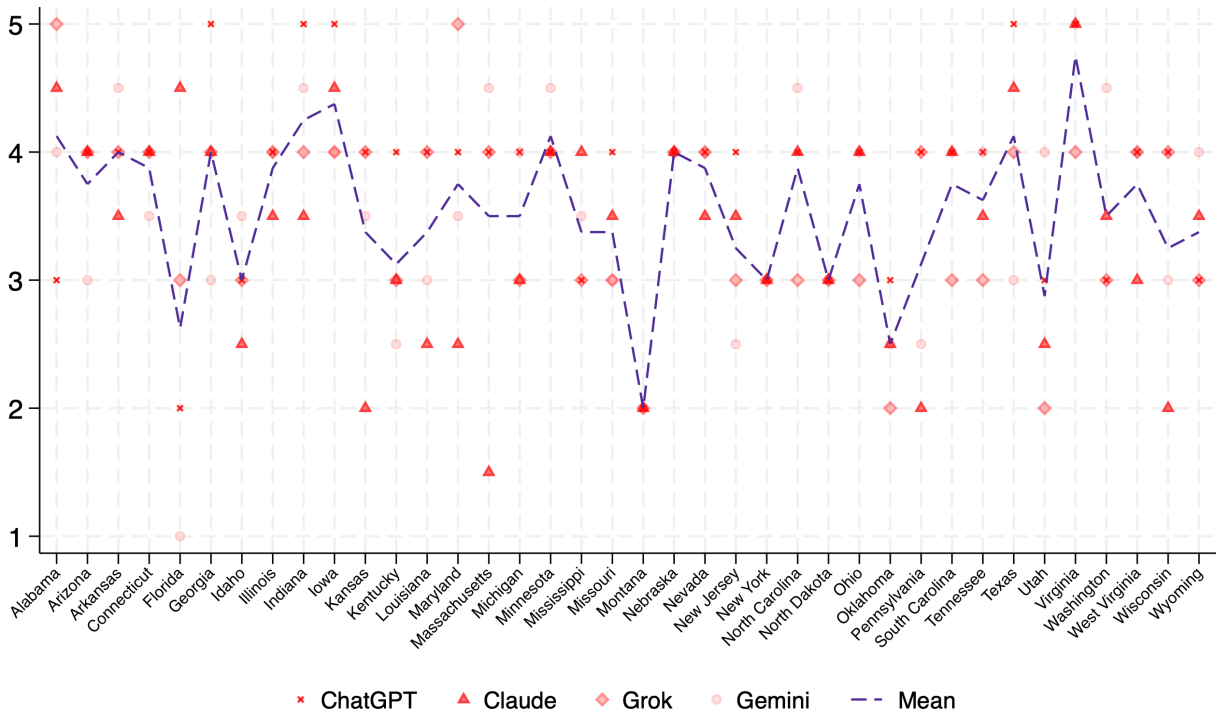


Figure A.4: This figure displays LLM-generated scores (ranging from 1 to 5) evaluating the scope and significance of state tax incentives. Scores produced by ChatGPT are marked with an x, those from Claude with a triangle, those from Grok with a square, and those from Gemini with a circle. The dashed blue line represents the average score across the four models for each state.

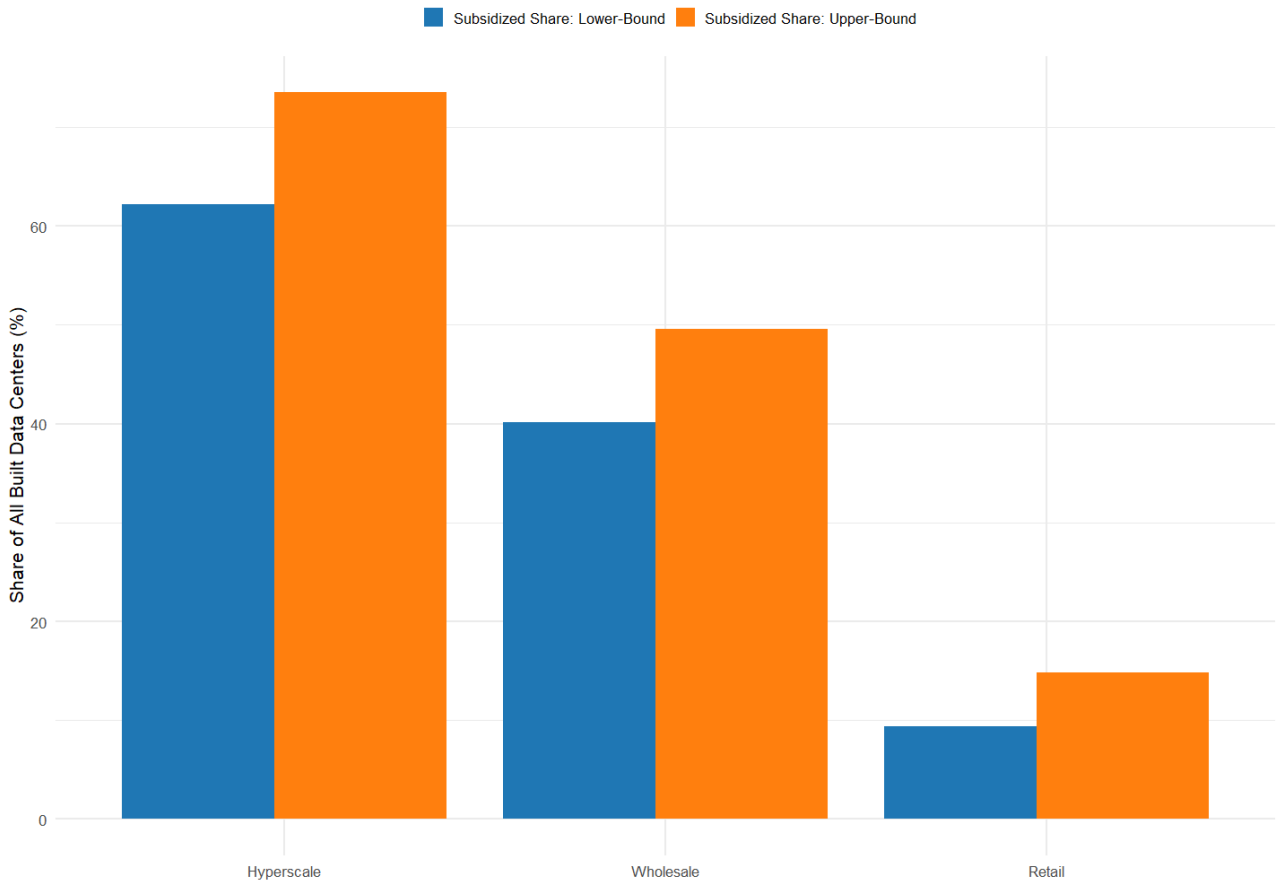


Figure A.5: This Figure displays the share of data centers built between 2000 and 2025 that likely qualified for state tax incentives, separately for the three categories of data centers: hyperscale, wholesale, and retail. The lower-bound estimates are based on the most restrictive size requirement in each state, and the upper-bound estimates on the least restrictive.

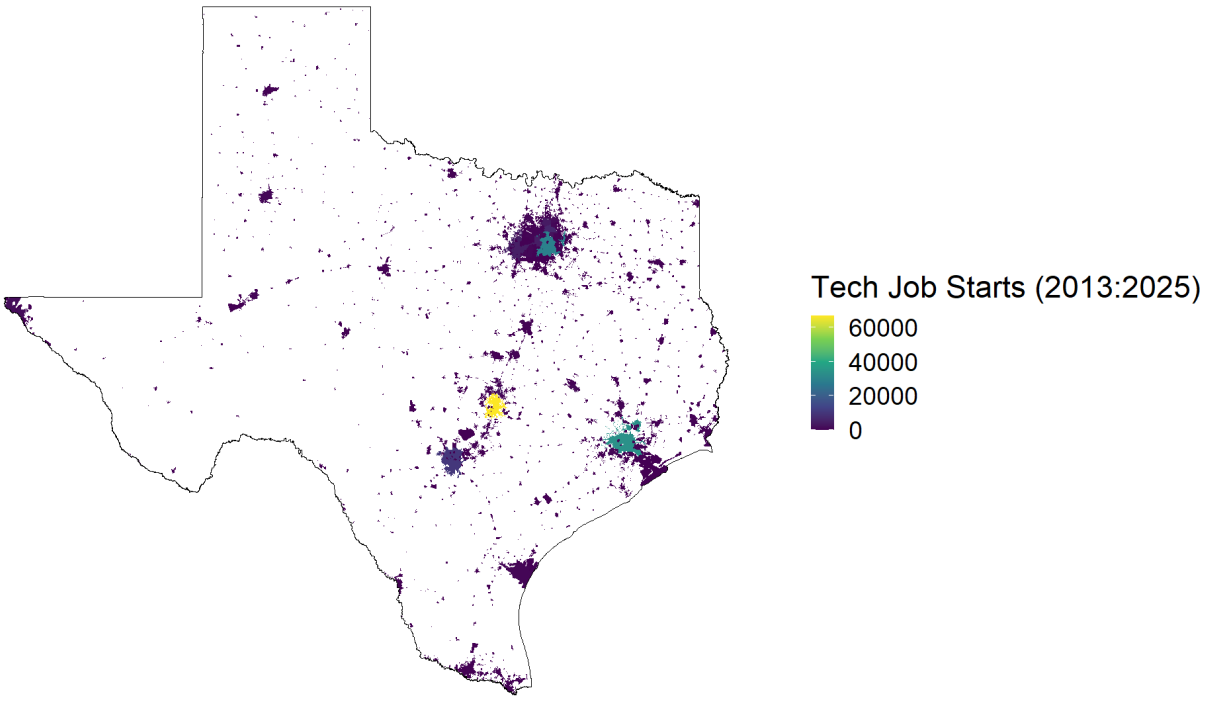
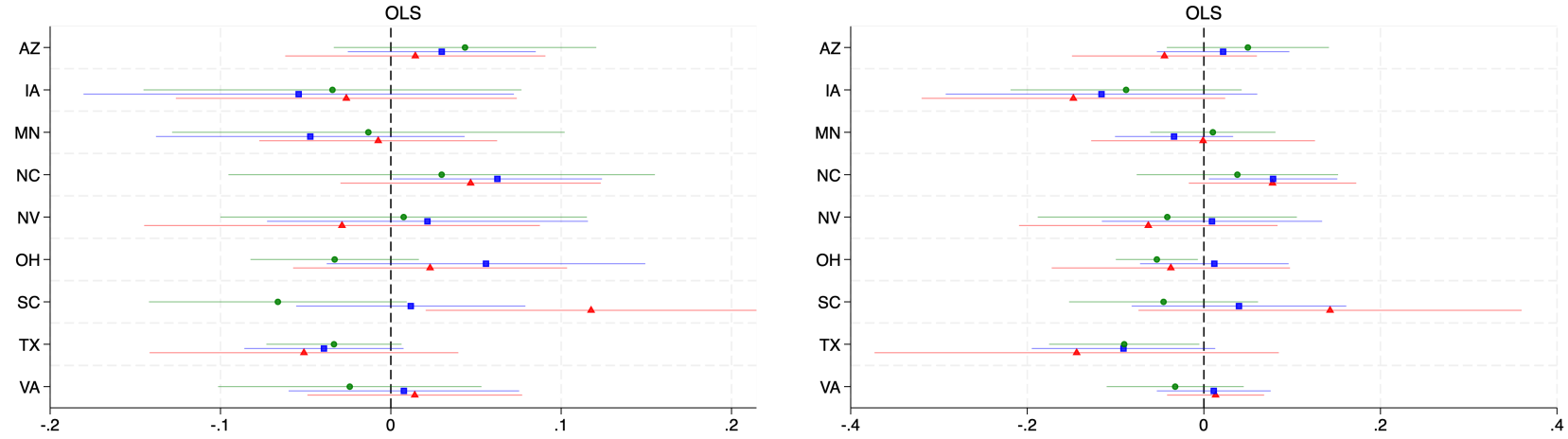
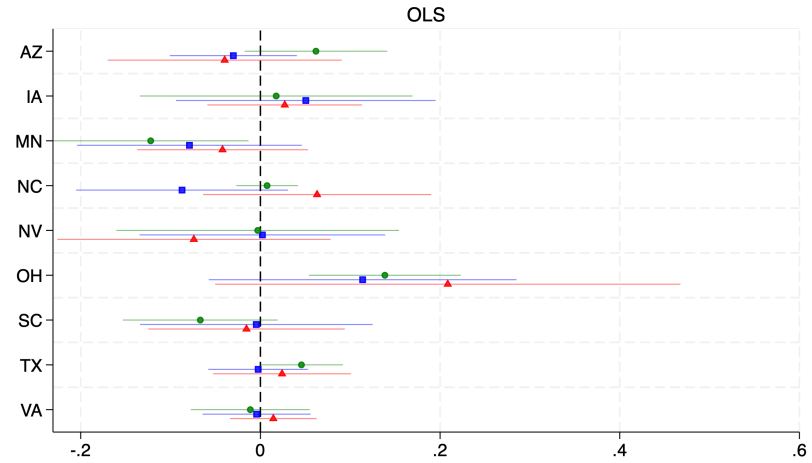


Figure A.6: This Figure shows the concentration of tech job starts by census place in Texas, using the total number of new tech job starts in the RevelioLab data over the period from 2013 to 2025. For this analysis, we classify individuals as starting a new tech job if they are employed by a firm in the tech sector.

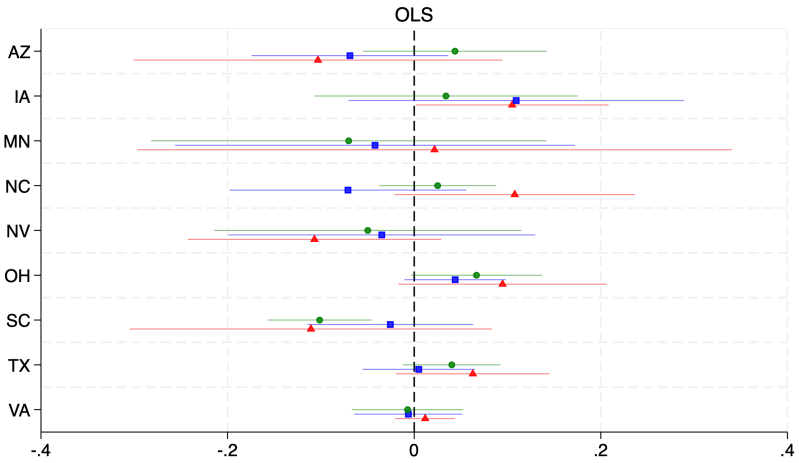


(a) 1-year ahead, 3 Matched Observations

(b) 1-year ahead, No matching



(c) 3-year ahead, 3 Matched Observations

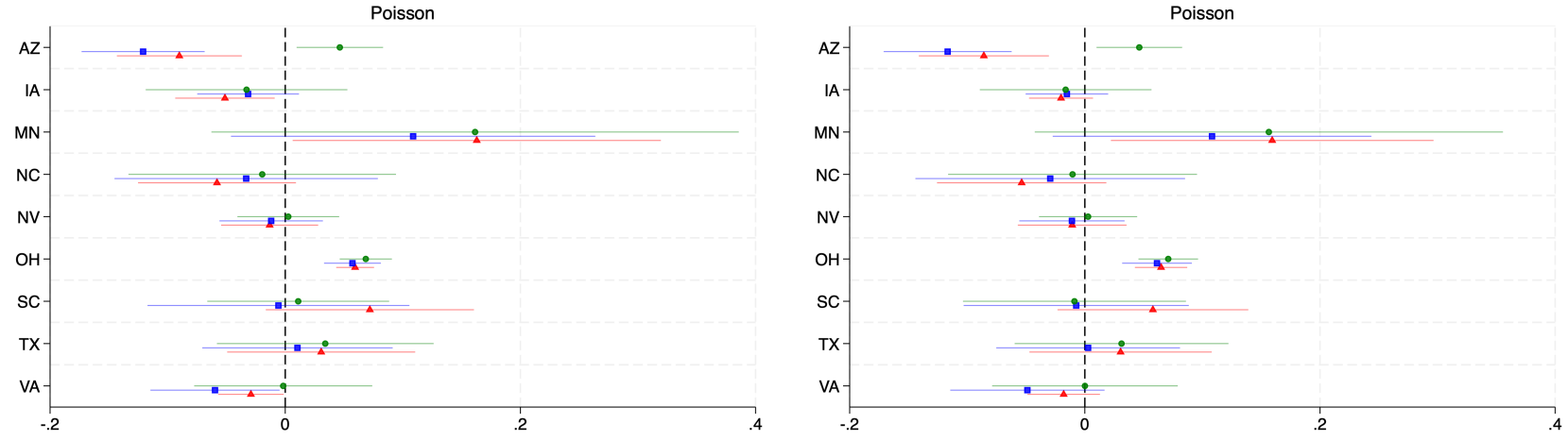


(d) 3-year ahead, No matching

Figure A.7: This Figure displays coefficient estimates and 95% confidence intervals from the following regressions

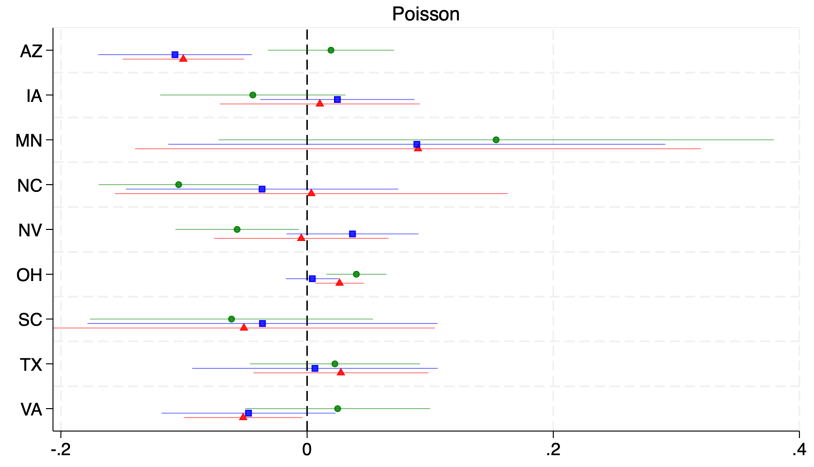
$$log(N_{TechJobs,c,t+h} + 1) = \beta z_{DCenter,c,t} + a_c + a_t + e_{c,t}$$

where $N_{TechJobs,c,t+h}$ is the number of new tech jobs in census place c and year $t+h$; $z_{DCenter,c,t}$ is a dummy equal to one if at least one new data center is completed within 10 miles of census place c in year t; a_c and a_t are census place and year fixed effects. Green circles represent estimates from specifications in which the number of new tech jobs is measured using profile description, blue squares represent estimates in which it is measured using standardized roles, and red circles represent estimates in which it is measured using employer characteristics. In the panels on the left, each place exposed to data center development is matched to three non-exposed places. In the panels on the right, there is no matching and all places within the state are included in the estimation sample. The horizon h is 1 year in the panels on the top row, and 3 years in the panels on the bottom row.

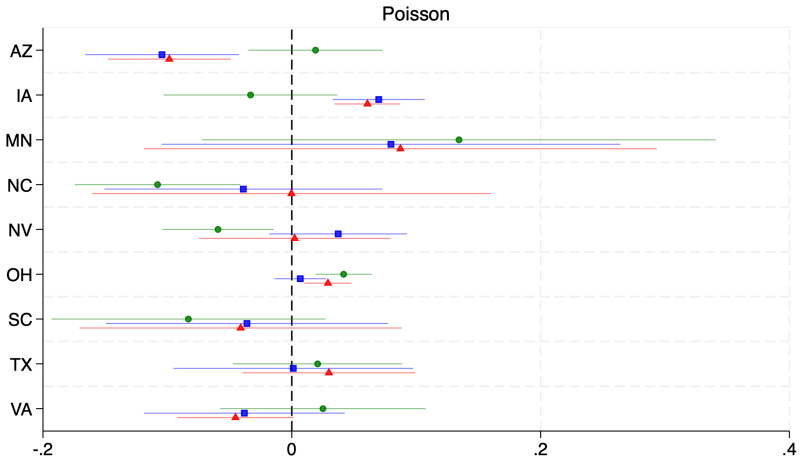


(a) 1-year ahead, 3 Matched Observations

(b) 1-year ahead, No Matching



(c) 3-year ahead, 3 Matched Observations



(d) 3-year ahead, No Matching

Figure A.8: This Figure displays coefficient estimates and 95% confidence intervals from the following regressions

$$N_{TechJobs,c,t+h} = \exp(\psi z_{DCenter,c,t} + p_c + p_t + u_{c,t})$$

where $N_{TechJobs,c,t+h}$ is the number of new tech jobs in census place or city c and year $t+h$; $z_{DCenter,c,t}$ is a dummy equal to one if at least one new data center is completed within 10 miles of census place c in year t; p_c and p_t are census place and year fixed effects. Green circles represent estimates from specifications in which the number of new tech jobs is measured using profile description, blue squares represent estimates in which it is measured using standardized roles, and red circles represent estimates in which it is measured using employer characteristics. In the panels on the left, each place exposed to data center development is matched to three non-exposed places. In the panels on the right, there is no matching and all places within the state are included in the estimation sample. The horizon h is 1 year in the panels on the top row, and 3 years in the panels on the bottom row.

Table A.1: Tax Incentives and Planned Data Center Development

	A: Hyperscale		
	$\mathbb{1}\{Built\}$	$ln(1+N)$	N
D	0.44***	0.93***	2.21***
	(3.22)	(3.75)	(3.02)
α	0.17	0.25	-0.54
	(1.52)	(1.51)	(-0.82)
R^2-adj	0.140	0.134	
N	50	50	50
	B: Wholesale		
	$\mathbb{1}\{Built\}$	$ln(1+N)$	N
D	0.19	0.50	1.11*
	(1.13)	(1.30)	(1.94)
α	0.42***	0.84**	1.20**
	(2.87)	(2.67)	(2.57)
R^2-adj	0.026	0.026	
N	50	50	50
	C: Retail		
	$\mathbb{1}\{Built\}$	$ln(1+N)$	N
$\mathbb{1}\{Incentive\}$	0.02	0.11	1.74
	(0.23)	(1.09)	(1.55)
α	0.08	0.06	-2.48**
	(1.02)	(1.02)	(-2.57)
R^2-adj	0.001	0.011	
N	50	50	50

This Table displays coefficient estimates (and associated t-stats) for the following regression equation:

$$y_s = \gamma D_s + \alpha + v_s$$

where y_s denotes a dummy equal to 1 if a new data center is planned in state s as of Q2 2025, the log of the number of new data centers planned, or the number of new data centers planned; D_s is a dummy variable equal to one if state s offers tax incentives in 2025. Estimates in panel A (B) (C) are for hyperscale (wholesale) (retail) data centers. T-stats are based on robust standard errors. Coefficients marked with ***, **, and * are significant at the 1%, 5%, and 10% level.

Table A.2: Model Calibration

Parameter	Description	Value	Source
$\tilde{\beta}_d$	Discount factor	0.91	Match industry research expected return
δ	Operating Income Growth	0.03	Industry research
C	Development Cost	\$11 M	Industry research
α	Revenue Effects of s_1,s_2,s_3	0.2	Match industry research cap rates
$DeprElEq$	Equipment Cost	\$0.495 M	Industry research
$DeprInfr$	Infrastructure Cost	\$0.138 M	Industry research
$ElCost_j$	Electricity Cost	State-Specific	EIA industrial electricity prices in 2025
$\mathbf{M}_j$	Mean Reversion Matrix	State-specific	Estimates from EIA and BLS data
$\mathbf{\Sigma}_j$	Residuals Matrix	State-specific	Estimates from EIA and BLS data
μ_j	Operating Income Constant	State-specific	Match investment shares
$\bar{\mu}$	Max Value of μ_j	\$0.18 M	$\approx$ 20% of mean income from $\mathbf{s}_j$ across states

This Table summarizes the model parameter calibration discussed in Section 4.3.

Table A.3: Model Results by State: Baseline Calibration Including Tax Incentives

State	PV ($M) $\mu_j = 0$	PV ($M)	V ($M) NPV	P_{inv}^{cond} (%)	Share (%) Constr.	Share (%) Constr. & Plan.
AZ	13.68	13.95	3.23	5.32	6.10	5.43
CA	-4.79	-4.79	-15.79	0.00	0.00	1.55
CO	12.01	10.35	-0.65	0.11	0.00	0.00
FL	13.00	10.67	-0.03	0.20	0.00	0.00
GA	13.46	13.54	2.74	3.25	1.22	3.36
IL	15.44	13.18	2.66	3.01	3.66	3.10
IN	11.25	14.05	3.40	6.28	2.44	6.46
IA	13.11	13.36	2.66	2.99	2.44	3.10
KS	14.41	11.63	0.95	0.54	0.00	0.00
KY	14.88	12.06	1.36	0.82	0.00	0.26
MD	-3.70	-3.70	-14.23	0.00	0.00	1.03
MA	8.11	8.04	-2.65	0.01	0.00	0.00
MI	13.30	10.78	0.24	0.27	0.00	0.00
MN	12.36	12.83	2.36	2.22	1.22	2.33
MO	13.76	12.72	1.93	1.45	0.00	1.55
NE	13.97	11.54	0.97	0.55	2.44	0.52
NV	13.11	13.02	2.36	2.22	4.88	2.33
NJ	6.06	6.04	-4.45	0.00	0.00	0.00
NY	14.31	11.65	0.85	0.49	1.22	0.26
NC	10.75	13.48	2.72	3.18	0.00	3.36
OH	13.96	14.96	4.41	17.21	20.73	17.31
OR	13.89	13.66	2.66	3.00	7.32	3.10
PA	12.57	13.36	2.66	2.99	2.44	3.10
SC	15.27	13.07	2.37	2.24	1.22	2.33
TN	15.77	12.89	2.24	1.97	1.22	0.52
TX	16.19	13.39	2.70	3.13	7.32	3.10
UT	13.83	11.47	0.94	0.54	0.00	0.52
VA	12.69	15.57	4.98	30.63	32.93	31.01
WA	16.16	13.28	2.78	3.40	0.00	2.33
WI	12.81	12.99	2.23	1.96	1.22	2.07

This Table displays model-implied new hyperscale data center valuations per UPS megawatt and conditional investment probabilities across states, as well as the share of new hyperscale data center projects in each state, out of all projects that are under construction, or all projects that are under construction and planned (across the 30 states) in 2025. The model is calibrated to account for the effects of state-level tax incentives. PV $\mu_j = 0$ is the present value of future operating income flows per megawatt after setting the constant component of income equal to zero (see equation 6). PV is the present value of future operating income. V is net present value (PV minus investment cost F). P_{inv}^{cond} is the probability of investing in a state, conditional on not delaying (see equation 7).

Table A.4: Regressions of μ_j on State-Level Characteristics

Variable	$\beta_{\mu 1}$	R^2	F-statistic	N. Obs.
ln(Total Population)	0.0443	0.0752	2.2754	30
ln(Population Density)	0.0417	0.1377	4.4704	30
ln(Median Income)	0.2958	0.1422	4.6415	30
Share Agriculture	-0.0414	0.0979	3.0381	30
Share Manufacturing	-0.1449	0.2200	7.895	30
Share Rural	-0.0010	0.0124	0.3518	30
Share HS Degree or Higher	0.0005	0.0001	0.0036	30
Share Bachelor Degree or Higher	0.0092	0.1324	4.2740	30
ln(Min Distance to Fiber Node)	-0.0002	0.0962	2.9817	30

This Table display coefficient estimates and statistics from the following regression equation:

$$\mu_j = \beta_{\mu 0} + \beta_{\mu 1} X_j + e_j,$$

where μ_j is the estimated fixed component of income in each state j, and X_j is a state characteristic. From the 2020 U.S. national census and the 2020 American Community Survey we calculate log total population, log population density, log median income, the share of individuals employed in manufacturing and agriculture, the share of individuals living in rural areas, and the share of household heads with high-school degree or higher, or bachelor degree or higher. For the log minimum distance from fiber nodes, we calculate the minimum distance across the state from the closest fiber node. To construct this measure, we collect the coordinates of the ten main U.S. fiber nodes.

Table A.5: Model Results by State: Counterfactual Calibration with No Tax Incentives

State	PV (\$M) $\mu_j = 0$	PV (\$M)	V (\$M) NPV	P_{inv}^{cond} (%)	Share (%) Constr.	Share (%) Constr. & Plan.
AZ	13.23	13.50	2.50	6.87	6.10	5.43
CA	-4.79	-4.79	-15.79	0.00	0.00	1.55
CO	12.01	10.35	-0.65	0.29	0.00	0.00
FL	11.77	9.44	-1.56	0.12	0.00	0.00
GA	13.14	13.22	2.22	5.17	1.22	3.36
IL	14.80	12.54	1.54	2.62	3.66	3.10
IN	9.69	12.49	1.49	2.50	2.44	6.46
IA	11.88	12.14	1.14	1.75	2.44	3.10
KS	13.88	11.10	0.10	0.62	0.00	0.00
KY	14.39	11.57	0.57	1.00	0.00	0.26
MD	-4.31	-4.31	-15.31	0.00	0.00	1.03
MA	7.60	7.53	-3.47	0.02	0.00	0.00
MI	12.68	10.16	-0.84	0.24	0.00	0.00
MN	11.65	12.12	1.12	1.72	1.22	2.33
MO	12.92	11.89	0.89	1.36	0.00	1.55
NE	13.40	10.98	-0.02	0.55	2.44	0.52
NV	12.56	12.47	1.47	2.44	4.88	2.33
NJ	5.37	5.36	-5.64	0.00	0.00	0.00
NY	13.99	11.33	0.33	0.78	1.22	0.26
NC	9.35	12.08	1.08	1.65	0.00	3.36
OH	13.37	14.37	3.37	16.35	20.73	17.31
OR	13.89	13.66	2.66	8.02	7.32	3.10
PA	12.09	12.88	1.88	3.67	2.44	3.10
SC	14.18	11.98	0.98	1.50	1.22	2.33
TN	14.54	11.66	0.66	1.09	1.22	0.52
TX	15.11	12.31	1.31	2.09	7.32	3.10
UT	13.20	10.84	-0.16	0.48	0.00	0.52
VA	12.14	15.03	4.03	31.54	32.93	31.01
WA	15.49	12.61	1.61	2.81	0.00	2.33
WI	12.41	12.58	1.58	2.73	1.22	2.07

This Table displays model-implied new hyperscale data center valuations per UPS megawatt and conditional investment probabilities across states, as well as the share of new hyperscale data center projects in each state, out of all projects that are under construction, or all projects that are under construction and planned (across the 30 states) in 2025. These results are for a counterfactual in which there are no tax incentives. PV $\mu_j = 0$ is the present value of future operating income flows per megawatt after setting the constant component of income equal to zero (see equation 6). PV is the present value of future operating income. V is net present value (PV minus investment cost F). P_{inv}^{cond} is the probability of investing in a state, conditional on not delaying (see equation 7).

Table A.6: Tech-Firm Industry Classifications

Code	Description
332510	Hardware Manufacturing
333242	Semiconductor Machinery Manufacturing
334111	Electronic Computer Manufacturing
334112	Computer Storage Device Manufacturing
334413	Semiconductor and Related Device Manufacturing
334416	Capacitor, Resistor, Coil, Transformer, and Other Inductor Manufacturing
334417	Electronic Connector Manufacturing
334419	Other Electronic Component Manufacturing
513210	Software Publishers
516210	Media Streaming Distribution Services, Social Networks, and Other Media Networks and Content Providers
517111	Wired Telecommunications Carriers
517112	Wireless Telecommunications Carriers (except Satellite)
517121	Telecommunications Resellers
517122	Agents for Wireless Telecommunications Services
517410	Satellite Telecommunications
517810	All Other Telecommunications
518210	Computing Infrastructure Providers, Data Processing, Web Hosting, and Related Services
519290	Web Search Portals and All Other Information Services
541330	Engineering Services
541511	Custom Computer Programming Services
541512	Computer Systems Design Services
541513	Computer Facilities Management Services
541519	Other Computer Related Services

This Table contains the list of NAICS codes used to identify tech firms in Section 5.

Table A.7: Balance of Matched Characteristics

	log(Income)	log(Home Value)	Age	Population	Poverty rate	Bachelor
AZ	0.11	0.08	2.82	0.99**	-5.52	3.86
	(1.10)	(0.51)	(0.70)	(2.50)	(-1.25)	(0.94)
IA	0.09	0.09	0.08	0.27	-1.41	4.11
	(1.29)	(1.12)	(0.05)	(0.60)	(-1.24)	(1.18)
MN	0.15*	0.31**	0.14	0.30	-1.11	8.97***
	(1.90)	(2.59)	(0.09)	(0.86)	(-0.86)	(3.17)
NC	0.02	0.05	-0.90	0.05	0.08	3.19
	(0.35)	(0.65)	(-0.67)	(0.11)	(0.04)	(1.06)
NV	0.15	0.11	-4.48*	2.01***	1.04	0.04
	(1.55)	(0.56)	(-1.77)	(3.00)	(0.41)	(0.01)
OH	0.08	0.08	0.33	-0.01	-1.18	4.81**
	(1.49)	(1.47)	(0.45)	(-0.03)	(-0.98)	(2.01)
SC	0.03	-0.05	-1.92	-0.10	-1.56	-2.27
	(0.25)	(-0.34)	(-0.76)	(-0.23)	(-0.54)	(-0.52)
TX	0.17***	0.21***	0.52	0.41**	-2.10**	7.59***
	(3.15)	(3.03)	(0.62)	(2.06)	(-1.98)	(3.37)
VA	0.42***	0.50***	-1.23	0.16	-3.63**	14.25***
	(5.86)	(6.13)	(-1.07)	(0.56)	(-2.56)	(4.77)

This Table displays estimates and corresponding t-stats for mean differences in matched demographic characteristics (measured using the 2012 American Community Survey) between census places exposed to data center development and places with no exposure. Each exposed place is matched to one non-exposed place. Exposure is defined as having at least one data center completed within 10 miles of the place boundary between 2013 and 2025. The analysis covers census places in Arizona, Iowa, Minnesota, North Carolina, Nevada, Ohio, South Carolina, Texas, and Virginia. Coefficients marked with ***, **, and * are significant at the 1%, 5%, and 10% level.

B State Incentive Scope Index

To construct the incentive scope index used in Section 3, we submit an identical prompt to four commercially available LLMs. In designing the prompt, we follow standard guidelines and current best practices. First, we specify a clear persona (a "Finance Professor working on an academic paper") and provide relevant context, including the topic of the query and the full list of states that have introduced tax incentives. Second, we describe the task (assigning a numerical score within a defined range) explicitly and outline the criteria, guidelines, and constraints that should govern the scoring.

To reduce the risk of hallucination, we instruct each model to articulate its reasoning and to prioritize accuracy. When available, we also activate "deep thinking" or "research" modes to encourage careful, step-by-step analysis. Finally, we clearly specify the desired output format and style: a column reporting the score and a corresponding column detailing the rationale for that score. The full prompt is reproduced below.

I'm a Finance Professor working on an academic paper examining whether and how state-level tax incentives have stimulated the development of data centers across the United States. I have attached a list of states that have implemented such incentives, along with the year of first implementation.

I need your assessment of the strength of incentives offered by each state. Please use a 1-to-5 scale, where 5 indicates very strong incentives and 1 indicates very weak incentives. Base your assessment on the following factors:

- *Type and duration of benefits: e.g., sales tax exemption, property tax exemption, electricity cost exemption, etc.*
- *Eligibility and qualifications: e.g., minimum investment, number of jobs created, average salary, etc.*

- *Other notable features of the regulation that may affect its attractiveness.*

Additionally, include a column explaining the rationale for your assessment. Prioritize accuracy and thoroughness over speed in your response.

C Assumptions for State Tax Incentive Estimates

Section 3.4. The analysis in this section relies on the minimum data center size required to qualify for state tax incentives. Although a few states specify this threshold in square feet, most define it as a minimum investment amount. To place all states on a common scale, we convert dollar-based thresholds into size-based thresholds using a set of simple assumptions. First, we translate the required investment into an implied square-foot size. For this conversion, we use the recent estimate in Guarino, Samandyk, and Webster (2025), which reports that the average development cost of a data center is approximately $1,500 per square foot. Thus, for state i, with minimum investment requirement D_i, the implied size threshold is $S_i^{2025} = D_i/1500$.

However, this threshold may be overstated because it relies on current construction costs, whereas many of the incentive laws were enacted well before 2025, when land, materials, and labor were considerably cheaper. Although we are working to obtain more detailed state-level construction cost indices, in the current version of the paper we adjust for price inflation using the CPI. Specifically, for each state i, we deflate the 2025 implied size threshold by the CPI inflation between the year the law was implemented and 2025. The adjusted threshold is therefore $S_i = S_i^{2025}/Inf_\tau$, where Inf_τ is calculated using the CPI index for urban consumers excluding food and energy, published by the Federal Reserve Bank of St. Louis.

It is also important to note that several states do not specify a single threshold but instead define multiple qualifying tiers. Lower thresholds often apply only if additional conditions are met (for example, if the data center is located in a low-income area). For the purposes of this section, we adopt the most restrictive threshold for each state.

Section 4.1. The calculations in this section rely on assumptions about which data centers are likely to have qualified for state incentives and the corresponding flow of tax benefits. To identify qualifying facilities, we combine the size thresholds derived in Section 3.4 with information on each data center's size and year of construction. If a facility was built after the implementation of the state's incentive program and meets the minimum size requirement, we

assume that it likely qualified for and received the associated incentives. Figure A.5 reports the share of data centers that qualify under both the highest and the lowest investment thresholds applicable in each state.

We then turn to calculating the incentives received by each data center. For this analysis, we restrict our focus to sales and use tax incentives, which constitute the most common form of subsidy, as shown in Table 2. Depending on the scope of each state's legislation, we compute up to three types of incentives: those related to electricity, equipment, and infrastructure. For example, for data centers in Indiana, we calculate exemptions for electricity and equipment, while in Minnesota we account for incentives covering equipment, infrastructure, and electricity, and so forth.

For electricity incentives, we estimate each data center's annual electricity consumption in 2025 by multiplying its utilized UPS power capacity by its power usage effectiveness ($PUE > 1$, reflecting efficiency losses). We then scale this hourly consumption by the number of hours in a year, assuming continuous 24/7 operation. To convert consumption into dollars, we multiply by the state industrial electricity price in 2023, adjusted for inflation to reflect 2025 dollars. Electricity prices are obtained from the Energy Information Administration. The annual sales tax incentive for electricity consumption is then computed as the product of the estimated annual electricity cost and the applicable state sales tax rate on electricity purchases.

For equipment incentives, we begin by estimating the total equipment cost for each data center. As in earlier calculations, we assume a development cost of $1,500 per square foot, an estimate based on an average power density of roughly 7,000 square feet per megawatt. We rescale this cost to reflect the actual power density of each facility and multiply by its size to obtain an estimated 2025 development cost. Following Guarino, Samandyk, and Webster (2025), we assume that equipment accounts for 38% of total development cost and depreciates at 10% per year. Accordingly, we assume that operators replace 10% of their equipment annually. The equipment sales tax incentive is then calculated by applying the relevant state sales tax rate to this estimated annual replacement cost.

For infrastructure incentives, we follow an analogous procedure. Infrastructure represents 21% of total development cost and is assumed to depreciate at 5% per year. We therefore estimate the annual infrastructure replacement cost and apply the state sales tax rate to compute the corresponding sales tax subsidy.