Digital Infrastructure and Local Welfare: Evidence from Data Center Openings*

Liu Ee Chia† Mingxuan Fan† Siyuan Hu‡ Qiang Wang§

January 14, 2026

Preliminary draft. Please do not cite without permission.
Most updated version here.

Abstract

This paper estimates the local welfare effects of hyperscale data centers as revealed through housing-market capitalization. We construct a panel dataset linking geocoded residential property transactions to the opening dates of hyperscale data centers and employ a difference-in-differences design that compares properties located within 5 km of a newly opened facility to properties located 20–50 km away. We find that hyperscale openings reduce nearby housing prices by approximately 6.8 percent on average, with a steep distance gradient: effects are largest within 1 km and dissipate to zero beyond 12 km. Consistent with hyperscale facilities being capital-intensive and labor-light, we find limited evidence of sustained increases in local employment or wages. In contrast, we document increases in residential electricity costs in host areas after commissioning. The results imply a localized welfare loss of \$3.6 billion for communities hosting hyperscale digital infrastructure.

JEL Codes: L86, Q51, R31

*We would like to thank Corbett Grainger, Daniel Phaneuf, Chris Timmins, XXX, and seminar participants at the University of Wisconsin-Madison for helpful comments. XXX provided excellent research assistance—all mistakes are our own.

†Department of Real Estate, NUS Business School, National University of Singapore. Emails: liueechia@nus.edu.sg (Liu Ee) and mfan@nus.edu.sg (Mingxuan).

‡Department of Agricultural and Applied Economics, University of Wisconsin-Madison. Email: shu248@wisc.edu.

§Haskayne School of Business, University of Calgary. Email: qiang.wang3@ucalgary.ca.

1 Introduction

Digital infrastructure has expanded rapidly across U.S. localities, driven by growth in cloud services and the increasing computational intensity of artificial intelligence workloads. Data centers represent the foundational infrastructure for digital services that individuals, businesses, and governments rely on every day. Projections indicate that by 2030, companies will invest nearly $7 trillion in capital expenditures on data center infrastructure globally.[1] A prominent feature of this expansion is the rise of hyperscale data centers—large, power-intensive facilities that concentrate high-performance computing equipment and require substantial supporting infrastructure. Local governments frequently court these facilities with tax incentives and zoning accommodations, motivated by the expectation that hyperscale investments will anchor local economic development. However, hyperscale data centers present a distinctive economic profile. Despite their scale and capital intensity, they generate minimal direct employment once operational, while potentially imposing localized costs through electricity-grid strain, water demand, and environmental externalities (Mytton (2021); Shehabi et al. (2024); Siddik et al. (2021)).

This paper provides national-level evidence on the net local welfare impacts of hyperscale data centers as revealed through housing-market capitalization. The relationship between data center development and local welfare is theoretically ambiguous. On the one hand, data centers have proven to be a lucrative business, contributing billions of dollars in tax revenue. Although they do not create substantial jobs, data centers may induce local investments that indirectly contribute to the labor market.[2] On the other hand, data center operations may impose local disamenities, which can adversely affect residents' quality of life. We ask whether proximity to hyperscale facilities is valued as a net amenity, reflecting fiscal spillovers and local demand increases, or instead as a net disamenity due to infrastructure costs and quality-of-life impacts.

To evaluate the net welfare effect, we examine housing-market capitalization as it captures homeowners' willingness-to-pay for localized net externalities. We combine proprietary data on data center locations, commissioning dates, and technical characteristics with nationwide residential transaction records and local economic outcomes. Our empirical strategy employs a stacked

[1] See, The cost of compute: A $7 trillion race to scale data centers.

[2] For example, in Virginia, large-scale data center development has coincided with increased investment in renewable energy infrastructure and the clustering of technology firms. See, 2024 Virginia Data Center Report.

difference-in-differences event-study design that exploits the openings of data centers. For each opening, we compare transactions within a near ring (within 5 km) to those within an outer ring ($20 - 50$ km) around the same facility, absorbing granular county-by-time shocks within each stacked cohort. The identifying assumption is that, absent commissioning, prices in the near and outer rings would have followed parallel trends conditional on fixed effects and transaction and property covariates.

We find that hyperscale facilities function as localized disamenities. In our preferred specification, transaction prices within 5 km fall by approximately 6.8 percent following the commissioning of a hyperscale facility. Event-time estimates show no detectable pre-trends and a decline that emerges within two years of opening and grows in magnitude several years later, consistent with capitalization as operations ramp up and local housing markets adjust. The negative capitalization effects exhibit a sharp distance gradient: the largest declines occur within 1 km, attenuate with distance, and become indistinguishable from zero beyond roughly 12 km. Moreover, the effects are concentrated around the largest facilities operated by major technology firms, suggesting that the scale and intensity of operations are key drivers of the observed housing market effects rather than data center entry per se.

We probe mechanisms to distinguish between offsetting demand-side benefits and localized cost channels. Using county-level employment and wage outcomes, we find no statistically meaningful evidence of sustained increases in local employment or wages following hyperscale openings across industries and income brackets. This pattern is consistent with hyperscale operations' being capital-intensive and highly automated, which limits labor-demand spillovers to the surrounding area. In contrast, we find evidence consistent with the pass-through of infrastructure costs. Using Census-tract measures of residential electricity costs, event-study estimates show a discrete upward shift (2%) after hyperscale opening that remains elevated for multiple years, while water-cost responses are smaller and less precisely estimated. Taken together, the evidence points to localized infrastructure and quality-of-life costs that are salient to nearby residents and capitalized into housing values.

The steep distance gradient in housing price declines, together with rising utility costs, suggests that quality-of-life deterioration is a primary driver of home devaluation near hyperscale data centers. Furthermore, the absence of an employment and income boom explains why there is a limited positive shock to local housing demand to offset the negative externalities associated with data center operations. While a potential increase in local tax revenue could translate into improved public services or lower taxes that benefit local homeowners, any such benefits appear too small to offset the direct adverse effects on home values in the short run.

These findings contribute to several strands of economics literature. First, we add to the hedonic valuation of localized industrial and infrastructure disamenities by treating hyperscale data centers as a distinct form of land use whose externalities arise from resource-intensive operations and supporting infrastructure. Previous work has documented negative housing capitalizations for a whole host of externalities in the industrial economy, including pollution (e.g., Bishop et al. (2020); Bui and Mayer (2003); Currie et al. (2015); Greenstone and Gallagher (2008)), electricity infrastructure (e.g., Davis (2011); Dröes and Koster (2016); Fraenkel et al. (2024); Hamilton and Schwann (1995); Heintzelman and Tuttle (2012); Hu et al. (2025); Zou (2020)), and resource development (e.g., Bartik et al. (2019); Gopalakrishnan and Klaiber (2014); Muehlenbachs et al. (2015)). We contribute to the emerging literature on the rapid development of digital infrastructure. Relatedly, we inform debates on the spatial incidence of place-based economic development policies (Allcott and Keniston (2018); Kline and Moretti (2014); Moretti (2010); Zheng et al. (2017)). While local governments often subsidize high-tech investments, our evidence suggests that, for labor-light digital infrastructure, localized costs borne by nearby residents can be substantial relative to observable local labor-market gains, raising distributional concerns for host communities.

We also contribute to the growing literature on the economic impacts of artificial intelligence. Prior work shows that AI investment is associated with higher firm growth (e.g., Babina et al. (2024); Chen et al. (2019); Hirvonen et al. (2022); Rock (2019)) and disruption to labor markets (e.g., Babina et al. (2023); Grennan and Michaely (2020); Abis and Veldkamp (2024); Cao et al. (2024); Acemoglu et al. (2022); Cockburn et al. (2018); Hirvonen et al. (2022)). The adoption of AI also reduces behavioral biases in trading decisions (e.g., D'Acunto et al. (2019)) and improves loan underwriting performance, but may increase disparities between minority and non-minority borrowers (e.g., Jansen et al. (2020); Fuster et al. (2022)). However, there is little evidence on the local welfare effects of data centers, which are the backbone of the AI industry. Feher et al. (2025) study the effects of data centers on local electricity prices, firm outcomes, and carbon dioxide emissions and find no detectable effects. We differ from those studies by focusing on hyperscale data centers, which are substantially larger and more power-intensive. Our results show that these facilities generate significant welfare losses for nearby homeowners.

The remainder of the paper proceeds as follows. Section 2 provides background on hyperscale data centers and potential channels of local impacts. Section 3 describes the data and sample construction. Section 4 outlines the empirical strategy. Section 5 presents the results, and Section 6 concludes.

2 Background

To understand why a facility ostensibly dedicated to the background processing of digital information might impact local housing markets, it is necessary to examine the physical and operational reality of the burgeoning U.S. digital infrastructure. The data center industry has evolved from small, on-premise server rooms located within corporate office buildings to massive, dedicated facilities. The hyperscale designation refers to a nascent class of data center architectures designed for massive scalability, cloud computing, big-data processing, and artificial intelligence workloads. Industry consensus loosely defines hyperscale data centers as ones with thousands of server units collocated in massive compounds. Physically, hyperscale data centers resemble massive distribution warehouses, often appearing as windowless, concrete boxes spanning tens to hundreds of thousands of square feet of floor space. In our empirical work, we operationalize hyperscale using facility classifications and power-capacity measures from our data, and we show in robustness analyses that the estimated effects are concentrated among the largest-capacity sites.

While the size and specifications of hyperscale data centers exhibit high variance, the defining characteristic is their voracious appetite for resources. Conducting complex cloud computing and training large language models requires clusters of high-performance Graphics Processing Units that run at much higher thermal densities than traditional Central Processing Units. Housing thousands of such clusters under one roof requires substantial redundant power. In practice, hyperscale data centers are distinguished from smaller data centers by the sheer scale of resources required for daily operation, with individual campuses now frequently exceeding 100 megawatts of power capacity (Shehabi et al., 2024). The additional burden on the local electricity grids necessitates the construction of dedicated high-voltage substations and transmission lines, which are themselves known disamenities associated with visual blight (Hamilton and Schwann, 1995). The costs of these upgrades can be socialized across the entire ratepayer base, leading to higher electricity bills for nearby residents. Siting data centers in capacity-constrained areas can also increase the risk of brownouts or necessitate procuring additional generation capacity and operating peaker plants, which can be dirtier and more expensive (Holland and Mansur, 2008).

Cooling requirements further concentrate externalities locally. Because the electrical energy consumed by servers is converted largely into heat, thermal management is a core operational constraint. While some newer facilities experiment with liquid immersion cooling, the industry standard for large-scale facilities remains evaporative cooling, which uses large cooling towers to evaporate water and reject heat into the atmosphere. A typical hyperscale data center can consume millions of gallons of water per day, equivalent to the supply of a small U.S. city (Mytton, 2021; Shehabi et al., 2024; Siddik et al., 2021). The high water demand places data centers in direct competition with residential and agricultural water needs. Visually, hyperscale data centers can produce large vapor plumes, which neighbors may perceive as industrial air pollution.

Another common, major complaint from residents living near data centers is noise pollution (e.g., Cary (2023)). Unlike the intermittent noise of traffic or airplanes, data center noise is constant, low-frequency, and often tonal. It emanates from the banks of industrial chillers, rooftop air-handling units, cooling-tower fans, and diesel backup generators, which are required to ensure extended uptime (Carlson and Teplitzky, 1974; Ellis, 1971; Zhouo et al., 2020). Low-frequency noise is particularly problematic because it can travel long distances with little attenuation and penetrate standard residential walls. Psychoacoustic research suggests that such constant, tonal noise is hazardous to human health (e.g., Basner et al. (2014); Broner and Leventhall (1983); Miedema and Vos (2003); Muzet (2007); Zou (2020)).

Despite these potential localized costs, municipalities routinely offer generous tax abatements and expedited permitting to attract hyperscale investments. These policies are typically justified by the prospect of job creation, local multipliers, and enhanced economic dynamism. However, hyperscale facilities are highly automated and employ relatively few workers once operational, which limits the scope for labor-market spillovers relative to traditional industrial anchors. As a result, the net local welfare effect is theoretically ambiguous: fiscal benefits and any broader agglomeration effects may be offset—potentially more than offset—by localized disamenities borne by nearby residents. The empirical analysis below evaluates this net effect using housing-market capitalization and complementary evidence on labor markets and utility costs.

3 Data

3.1 Data Center Openings Database

We obtain detailed facility-level information on U.S. data centers from the S&P Global 451 Research Global Database Knowledge Base. The database records a wide range of characteristics for each facility, including geocoordinates, construction year and/or operational year, data center type (e.g., retail, colocation, hyperscale), power capacity and utilization, operational and utilized floor space, and the operating company and facility owner. Figure 1 maps data center construction over the years across the contiguous United States. Our final sample includes all data centers with non-missing geocoordinates and a non-missing construction year or operational year. We define the treatment event as the facility commissioning (operational) year whenever available; when commissioning dates are missing, we use construction year as the closest proxy.

3.2 Residential Property Transactions

We obtain housing transaction data from CoreLogic, a widely used commercial database of U.S. property sales compiled from public deed and assessor sources. The data contain property-level sale records, including transaction price and date, along with a rich set of transaction and property characteristics. We apply several filters. First, we restrict the sample to residential properties (single-family homes, condominiums, duplexes, and apartments). Second, we retain arm's-length transactions and exclude intra-family transfers. We study transactions from 2000–2025 and construct inflation-adjusted sale prices expressed in 2010 dollars. To estimate the effects of data center proximity, we construct a stacked sample around each data center. For each facility, we retain transactions within a 50 km radius, classify properties within 5 km as treated, and use properties 20–50 km away as controls to limit potential spillovers in the baseline specification; in robustness checks, we reintroduce this intermediate band and vary the distance cutoffs.

We repeat this procedure for each data center and stack the resulting subsamples into a single estimation dataset. To address overlapping catchments and clustered facilities, we impose two additional restrictions. First, if a property is exposed to multiple data centers over time, treatment is assigned based on the first exposure. Second, we exclude from the control group any transactions occurring within 20 km of any data center, ensuring that control observations are not plausibly affected by nearby facilities. The final estimation sample stacks all subsamples after applying these filters.

3.3 Census and Other Data

We obtain local community characteristics at the census tract level to examine heterogeneity and explore potential mechanisms. Using the American Community Survey (ACS) 5-year estimates (2006–2019) and the 2000 and 2010 Decennial Censuses, we compile tract-level demographic and socioeconomic measures, including population, income, racial composition, and local utility prices. To assess local economic impacts, we draw on two sources of employment and wage data. First, the Longitudinal Employer–Household Dynamics (LEHD) program's Origin–Destination Employment Statistics (LODES) provide annual employment counts by workplace location (census tract) and industry, allowing us to measure overall employment as well as data-center-related employment in each tract. As a robustness check, we aggregate employment outcomes to the county level and use the Bureau of Labor Statistics (BLS) Quarterly Census of Employment and Wages (QCEW) as an alternative county-level source. From QCEW, we extract total covered employment and average weekly wages by county-year across industry sectors.

We collect water utility service-area boundaries from SimpleLab's TEMM (Tiered Explicit, Match, and Model) dataset, which compiles a nationwide layer of community water-system geographies by assigning each system the highest-fidelity boundary available: (i) explicit polygons from state or utility sources when available; (ii) matched U.S. Census place polygons when an acceptable one-to-one match exists; and (iii) otherwise, modeled circular service areas centered on the best-available system centroid using a statistical model trained on labeled systems. We link TEMM to the housing transactions via a spatial join: each geocoded transaction is assigned the water-system boundary that contains its location, with unmatched transactions flagged accordingly.

To characterize the seasonal directionality of potential data-center externalities, we use near-surface wind vectors from the North American Regional Reanalysis (NARR), obtained from the U.S. National Oceanic and Atmospheric Administration's Physical Sciences Laboratory. The NARR provides gridded long-term monthly mean winds at a height of 10 meters over North America, reported as horizontal and vertical components in meters per second. For each hyperscale data center site, we spatially assign the two speed measurements at the facility location and produce an azimuth (degrees clockwise from true north) indicating the direction toward which the air is moving. We then compute the great-circle bearing from each data center to each parcel centroid and construct monthly directional exposure indicators. For each calendar month, a parcel is classified as downwind if the absolute angular difference between the parcel bearing and the site-specific wind azimuth is within ± 30 degrees (i.e., a 60-degree circular sector), and vice versa for the upwind indicator. This procedure yields 24 monthly downwind and upwind indicators for each parcel, which we use to test whether estimated treatment effects concentrate in wind-favored directions consistent with propagation and transport mechanisms of environmental elements.

3.4 Summary Statistics

Table 1 reports transaction-level means for sales within 5 km of a data center site and for sales 20–50 km away, separately for (i) all data centers and (ii) hyperscale data centers. When focusing on all data centers across the United States, transactions near sites have higher mean sale prices, but involve smaller homes and substantially older housing stock and are located in neighborhoods with a higher Black share, lower median income, and higher unemployment. For properties near hyperscale facilities, by contrast, mean prices are slightly lower in the near ring, while observable housing and neighborhood characteristics are broadly similar across distance bands. Overall, these summary statistics highlight the importance of the fixed effects and rich controls in our baseline specification.

4 Empirical Framework

Our primary objective is to estimate the causal effect of data center openings on local housing markets. A central identification challenge is that facility locations are not randomly assigned: developers may select sites based on pre-existing infrastructure, growth prospects, or political-economy considerations that are correlated with underlying housing-market trends. We address this concern using a stacked difference-in-differences design implemented in an event-study framework, with the unit of observation being a residential property transaction for property i in year t.

Specifically, we construct a stacked dataset in which each event—defined as the opening of a particular data center—forms a separate cohort of housing transactions. Within each cohort, we classify observations as treated or control based on distance to the opening site. In the baseline specification, treated observations are transactions within a narrow radius of the data center ($\leq 5km$). Control observations are transactions located 20–50 km from the same data center facility, restricted to the same county and the same transaction year–month as the treated transactions. This control-group construction is intended to capture transactions exposed to similar county-level demographic, economic, and housing-market conditions, while remaining sufficiently distant to avoid the facility's immediate disamenity zone.[3] The identifying assumption is that, absent data center opening, housing prices in the near and outer rings would have followed parallel trends after conditioning on fixed effects and observable transaction and property characteristics.

Our baseline specification is specified as follows:

$$y_{i,t,j,c,d} = \beta_1 \cdot \mathbf{1}\{\text{Dist}_{id} \leq 5km\} \times \mathbf{1}\{\text{PostOpening}_{t,d} = 1\} + X'_{i,t}\Gamma + \alpha_{j,d} + \theta_{c,t,d} + \epsilon_{i,t,j,c,d}, \qquad (1)$$

where $y_{i,t,j,c,d}$ is the natural log transaction price of property i, located in census tract j and county c, transacted at year-month t, in stacked cohort d (defined by a specific data-center opening). $\text{Dist}_{i,d}$ is the straight-line distance from the property to data center d, and $\text{PostOpening}_{t,d}$ equals to one for transactions in periods after the facility d is commissioned. $X_{i,t}$ includes a rich set of property and transaction characteristics, $\alpha_{j,d}$ are tract fixed effects, and $\theta_{c,t,d}$ are county-by-time fixed effects; all fixed effects are interacted with data center cohort d to reflect the stacked design. The coefficient of interest, β_1, captures the average differential change in prices for properties within 5 km relative to the comparison group before versus after the opening of data center d, and can be interpreted as the capitalization of the net local amenity or disamenity associated with the facility.

We also estimate an event-study version of Equation (1) by replacing the post-opening indicator with a set of relative-time indicators:

[3] We recognize that the choice of distance bands is inherently somewhat arbitrary. Accordingly, in later sections we examine distance-gradient effects by estimating separate regressions for a series of finer, non-overlapping distance bins (e.g., 0–1 km, 1–2 km, . . . , up to 19–20 km). In the robustness section, we also vary the distance bands to ensure consistency of results.

$$y_{i,t,j,c,d} = \sum_{\ell=-K,\ell\neq-1}^{K} \beta_\ell \mathbf{1}(\text{Dist}_{i,d} \leq 5km) \times \mathbf{1}(\text{YearsSinceOpen}_{i,t} = \ell) + X'_{i,t}\Gamma + \alpha_{j,d} + \theta_{c,t,d} + \epsilon_{i,t,j,c,d}. \tag{2}$$

where $\mathbf{1}(\text{YearsSinceOpen}_{i,t} = \ell)$ indicates that transaction i occurs ℓ years relative to the opening of data center d, and $\ell = -1$ is omitted as the reference period. We include a full set of leads and lags from $\ell = -25$ to $\ell = 22$ of data center opening, while maintaining the same covariates and fixed effects as in the baseline model. To improve precision at long horizons, we bin extreme event times by pooling all $\ell \leq -5$ into a single pre-period bin and all $\ell \geq 5$ into a single post-period bin. β_{-1}, the year immediately before opening, is omitted as the reference, so all estimates are interpreted relative to the pre-opening year. This event-study specification both tests for differential pre-trends and traces the dynamic evolution of capitalization effects after commissioning.

5 Results

5.1 Impact on Local Housing Markets

Table 2 reports pooled post-opening capitalization effects estimated in a stacked difference-in-differences design that compares housing transactions within 5 km of a given data center (treated group) to transactions in a 20–50 km ring around the same facility (control group). Columns progressively add covariates and fixed effects to address selection in siting and contemporaneous shocks. Panel A pools all data center facilities in the sample, while Panel B restricts attention to hyperscale data centers.

In both panels, Column (1) presents a parsimonious specification that includes Census-tract fixed effects and transaction year-by-month fixed effects. Tract fixed effects absorb time-invariant neighborhood characteristics such as baseline amenity levels, long-run development patterns, while year-by-month fixed effects control for nationwide housing-market conditions and seasonality. Column (2) adds a rich set of property and transaction controls, including structural attributes (e.g., type, size, age, rooms) and sale characteristics (e.g., arms-length indicators and financing). These covariates address concerns that transaction composition differs systematically between the near and outer rings within a cohort. In addition, we include flexible distance-to-facility fixed effects (50 one-kilometer bins) to absorb smooth spatial gradients in prices around the site that are unrelated to the opening itself (e.g., proximity to highways, employment centers, or industrial corridors).

Column (3) instead augments Column (1) with transaction-level controls, including indicators for cash purchases, resales, short sales, and foreclosure/REO transactions; buyer and seller characteristics (e.g., corporate status, local status); and additional controls capturing prior sale circumstances and buyer–seller relationship indicators where available. These variables proxy for variation in liquidity, distress, and investor activity that may covary with local market conditions and could otherwise confound treatment effects. Column (4) includes both the property and transaction controls, providing a more saturated specification that jointly accounts for compositional differences in transacted housing stock and shifts in transaction types.

Columns (5) and (6) strengthen identification by absorbing increasingly localized time-varying shocks. Column (5) replaces the common year-by-month fixed effects with state-by-year-by-month fixed effects, restricting identification to within-state variation and controlling for state-level housing and non-housing shocks (e.g., policy changes, statewide labor-market conditions, or credit-market shifts). Column (6) further replaces these with county-by-year-by-month fixed effects, thereby netting out any county-level shocks that could jointly affect data center openings and housing prices—such as changes in local infrastructure investment, zoning and permitting environments, or countywide adjustments in utility rates and service provision. In this specification, identification comes from comparing treated and control transactions within the same county and month, at similar baseline distances, before versus after the opening.

Across specifications, the estimates are stable in both sign and magnitude. Panel A shows economically small and statistically muted average effects when pooling all data center facilities.[4] By contrast, Panel B focuses on hyperscale openings. In the preferred specification in Column (6)—which combines rich property and transaction controls with county-by-year-by-month fixed effects—prices for transactions within 5 km fall by 6.8% relative to the 20–50 km ring after a hyperscale opening. The persistence of negative effects as we add high-dimensional fixed effects and increasingly granular controls suggests that the results are not driven by coincident state- or

[4] One possible explanation for the muted effects when pooling all data centers is that smaller facilities generate more localized spillovers, so a 5 km treatment radius may be too broad and may attenuate estimated impacts. To assess this possibility, Appendix Table A-1 restricts the sample to non-hyperscale facilities and redefines treatment using narrower distance bands (e.g., within 1 km, 1–2 km, and so on). The estimates remain economically small and statistically indistinguishable from zero across these specifications.

county-level housing cycles, nor by compositional shifts in the types of homes transacting near openings. Overall, the contrast between Panels A and B implies that localized disamenities are concentrated among hyperscale facilities. This distinction is consequential given the ongoing expansion of hyperscale capacity globally, indicating that distributional effects on nearby homeowners may depend more on facility scale than on the presence of a data center per se.

To assess dynamic responses and probe the identifying parallel-trends assumption, we next estimate and plot event-study coefficients. Figure 2 reports the estimated effects of hyperscale data center openings on housing prices by event time, normalized to the year immediately prior to opening. Pre-opening coefficients are close to zero and statistically indistinguishable from zero over all the years preceding commissioning, providing no evidence of differential pre-trends between the treated and control rings. After commissioning, the coefficients become negative and statistically different from zero by year 2. The decline deepens through year 4—reaching approximately -13%—and then partially attenuates, stabilizing at roughly -5% thereafter. Overall, these dynamics indicate a discrete post-opening decline in nearby housing prices with limited evidence of anticipatory adjustment, lending support to the comparability of treated and control areas and reinforcing a causal interpretation of the stacked difference-in-differences estimates.

5.2 Distance Gradient in Price Capitalization

To assess how capitalization varies with proximity and to evaluate whether our baseline 5 km treatment definition masks important heterogeneity, we re-estimate the stacked difference-in-differences specification allowing the post-opening effect to vary flexibly across non-overlapping distance bins (0–1 km, 1–2 km, ..., 19–20 km), using transactions in the 20–50 km ring as the reference group. This exercise serves two purposes. First, it provides a more transparent characterization of the spatial incidence of data center openings and helps identify the geographic scope of any externality zone. Second, it offers an additional diagnostic for mechanisms: channels that operate primarily at broader administrative or market levels (e.g., countywide housing demand shifts, local public-service provision, or utility-rate changes that apply uniformly within a service territory) would not typically generate sharp attenuation over a few kilometers.

Figure 3 reveals a steep distance gradient in housing-price responses. Homes within the first kilometer experience the largest devaluation, approximately 15%.[5] The effect attenuates to roughly 9% over the 1–4 km range and remains economically meaningful at around 5% through 4–12 km, before declining further and becoming statistically indistinguishable from zero at larger distances. This spatial pattern is consistent with highly localized costs that are strongest near the facility and decay with distance, reinforcing the interpretation that the average 0–5 km estimate reflects an aggregation of substantially larger impacts very close to the site. Substantively, the gradient highlights that the distributional burden of hyperscale openings is concentrated among nearby homeowners, which is directly relevant for siting decisions, buffer-zone design, and the evaluation of place-based incentives that may generate diffuse fiscal benefits but localized housing-market costs (e.g., Hamilton and Schwann (1995); Zou (2020)).

5.3 Robustness Checks

Alternative Staggered DD Estimators. Figure 4 assesses whether the baseline results are sensitive to the particular staggered-adoption estimator or specification used, namely the stacked difference-in-differences approach. The figure compares dynamic treatment effects obtained from several staggered DiD approaches that are designed to be robust to the negative-weight and treatment-effect-heterogeneity issues that can bias two-way fixed-effects estimators in settings with staggered timing ((Borusyak et al. (2024); Callaway and Sant'Anna (2021); De Chaisemartin and d'Haultfoeuille (2024); Sun and Abraham (2021))). Across methods, the qualitative conclusion is consistent and robust. Estimated coefficients in the pre-opening window are tightly centered around zero, with no systematic evidence of anticipatory declines prior to the operational date, supporting the plausibility of parallel trends conditional on the fixed effects and controls. In the post-opening period, most estimators indicate a decline in nearby house prices, with effects that grow in magnitude over longer horizons. Although point estimates differ across procedures—for example, the Sun and Abraham estimator yields larger effects—the dynamic pattern is similar: modest declines in the first few years after opening, followed by larger negative effects several years later.

Alternative Control Groups. The baseline specification defines treatment as transactions

[5] The corresponding standard error is relatively large due to the small number of housing transactions observed within this narrow distance band around hyperscale facilities (slightly above 2000 transactions).

within 5 km of a hyperscale opening and excludes an intermediate 5–20 km buffer ring from the comparison group to mitigate spillovers. Because these distance cutoffs are necessarily somewhat ad hoc, we conduct a set of robustness checks that vary both the treatment and comparison definitions. Table 3 reports the results.

First, to assess sensitivity to the buffer-ring exclusion, Column (1) removes the buffer and uses all transactions in the 5–50 km ring as controls. The estimated price decline falls to 3.4%, smaller than the baseline 6.8%, consistent with spillovers extending beyond 5 km and attenuating through roughly 15 km (Figure 3). Motivated by the distance-gradient evidence, Columns (2) and (3) instead use comparison rings of 15–50 km and 15–30 km, resulting in estimate price declines of 5.6% and 5.3%, respectively. When the excluded ring is expanded by using 30–50 km as the comparison group, the estimated decline increases to 8.9%, again align with attenuation from spillovers when controls are drawn from closer distances.

We also consider comparison groups that do not rely solely on distance rings. Column (5) uses "later-treated" locations—transactions within 5 km of facilities that open later in the same county—as controls, addressing concerns that neighborhoods near data centers may differ systematically from those farther away. This specification yields an estimated price decline of 4.7%. Finally, Column (6) defines exposure at the census-tract level, comparing tracts that ever host a hyperscale opening (dated to the first opening) to neighboring never-treated tracts. We find a smaller estimated effect of −1.6%. Overall, the negative price response is robust, and the systematic attenuation under broader comparison rings or coarser exposure definitions reinforces the interpretation that impacts are concentrated very near the hyperscale facilities, providing implications for siting policy and for evaluating distributional consequences for nearby homeowners.

Alternative Definitions of Hyperscale Data Centers. In the baseline analysis, we classify hyperscale data centers using the facility-type indicator in the raw dataset, which distinguishes hyperscale facilities from wholesale, retail, and other categories. Here, we consider an alternative, scale-based definition using the dataset's measure of Total Uninterruptible Power Supply (UPS) power.[6,7] Table 4 reports estimates in which the treated data centers included in the estimation

[6] Total UPS power measures the cumulative installed, usable UPS capacity available at the rack for client IT equipment. Because UPS capacity must be provisioned to backstop the critical compute load—and scales with the number of racks/servers and the facility's engineered power density—it provides a practical proxy for data center scale. (i.e., "IT load" supported by the facility).

[7] Industry and policy sources commonly characterize hyperscale facilities as supporting at least 100 MW of power

sample are defined by progressively lower UPS-power thresholds across columns.

The results exhibit pronounced scale dependence. Restricting the sample to facilities with at least 100 MW of UPS power yields large and precisely estimated post-opening effects: nearby housing prices decline by 13.5% (Column 1). Lowering the threshold to 80 MW produces a smaller but still statistically significant decline of roughly 7% (Column 2). As the threshold is reduced further (60, 40, and 20 MW; Columns 3–5), the estimates become economically small and statistically indistinguishable from zero. Overall, the pattern suggests that the effects are concentrated among the largest facilities.

5.4 Mechanisms

We next examine complementary outcomes to study the channels underlying the housing-market response. The central challenge in evaluating hyperscale data center operations is the potential for localized disamenities to be generated alongside any economic benefits. We focus on labor-market outcomes, residential utility costs, and environmental externalities—dimensions that feature prominently in local policy debates over data center siting and incentives.

5.4.1 The Labor Market Effect

A common justification for state and local incentives is that attracting data centers will stimulate local economic activity—through ongoing operational activity, investment, and some permanent employment—thereby expanding the local tax base and potentially increasing housing demand and offsetting any localized disamenities. A contrasting view emphasizes that these facilities are highly automated and, once constructed, require relatively few workers to operate. We therefore examine local labor-market responses to hyperscale data center openings directly.

Using Census-block–level LODES Workplace Area Characteristics (WAC) data, we estimate event-study specifications for employment by industry, comparing blocks within 5 km of a hyperscale opening (treated) to blocks 20–50 km away (controls) in a stacked difference-in-differences setting, with distances computed using block centroids (Figure 5). The block-level estimates suggest limited employment responses overall. In sectors most plausibly linked to data center operations—Utilities, Information, and Professional, Scientific, and Technical Services—we do not find

capacity (Offutt and Zhu (2025))

a sustained increase in employment in the first few years after facility commissioning. At longer horizons, we observe at most a modest increase in Professional, Scientific, and Technical Services of approximately 4%, while estimates for Utilities and Information remain close to zero. Consistent with these patterns, we find little evidence of differential changes in employment across earnings bins (e.g., below versus above $3,333 per month).

Employment effects may also occur through complementary activity that need not be concentrated in the immediate vicinity of the facility—for example, related suppliers, contractors, or downstream users that locate elsewhere within the region. To capture such broader spillovers, we complement the block-level analysis with county-level estimates that aggregate employment across Census blocks within each county. In this specification, treated units are counties hosting a hyperscale data center, and the comparison group consists of later-treated counties within the same state.[8] Figure 6 illustrates the county-level event-study estimates by industry: coefficients fluctuate around zero and confidence intervals are wide, with no persistent break at the opening date.[9] We further corroborate these findings using county-level outcomes from the QCEW, including establishment counts, employment, total wages, and average weekly wages; the corresponding event-study estimates likewise show muted effects across these measures, both when pooling all industries and when restricting attention to the Information sector (Appendix Figure A-2). Overall, the labor-market evidence neither points to a systematic change in employment nor a reallocation across earnings groups that could plausibly explain the decline in housing prices documented in Section 5.1.

5.4.2 Utility Costs

Hyperscale facilities add large, geographically concentrated electricity loads that can trigger transmission upgrades, capacity procurement, and other grid investments. Because a portion of these costs may be recovered through retail rates, hyperscale openings could increase electricity prices faced by households. We examine this channel using Census-tract measures of average annual

[8] Using later-treated counties as controls leverages variation in treatment timing within the set of ultimately treated counties, which is less vulnerable to selection on long-run growth prospects than comparisons to never-treated counties.

[9] As an additional robustness check, we aggregate block-level employment to commuting zones, following the literature that treats commuting zones as integrated local labor markets that capture commuting-based spillovers beyond county boundaries. As shown in Appendix Figure A-1, the results are qualitatively unchanged, with employment responses remaining muted after hyperscale facilities openings.

household electricity and water costs constructed from ACS data. Figure 7 reports event-study estimates comparing tracts hosting hyperscale facilities to surrounding non-treated tracts.[10] Two patterns emerge. First, electricity costs exhibit a clear upward shift after commissioning: post-treatment coefficients turn positive and remain elevated ($\approx 2\%$) for several years, while pre-period coefficients are comparatively flat and statistically indistinguishable from zero. This pattern is consistent with industry accounts that large data centers can raise procurement costs and necessitate grid investments that are subsequently passed through to households (e.g., Blunt and Hiller (2026); Wade et al. (2025)).

Second, water-cost responses are economically small and statistically indistinguishable from zero, with no sustained post-opening increase. However, a prominent concern is water stress: hyperscale data centers may withdraw substantial volumes of water for cooling, potentially tightening local water availability even if household water bills do not adjust in the short run as measured in the ACS. To probe whether the housing-price effect is sensitive to water-system-level shocks, we exploit water-service boundary data from SimpleLab TEMM and estimate specifications that flexibly absorb water-utility variation.

In Table 5, we estimate specifications that flexibly absorb water-utility shocks by interacting water-system boundaries with: 1) year fixed effects and 2) county-by-year fixed effects in Columns (1) and (2), respectively. Relative to our preferred baseline estimate of 6.8%, the price impacts attenuate sharply to $0.9 - 1.4\%$ and become statistically insignificant. To rule out the possibility that the attenuation reflects a reduced estimation sample size after matching transactions to water-service boundaries, Columns (3) and (4) retain unmatched transactions by assigning a missing-category indicator for the boundary identifier. The resulting coefficients of 0.032 and 0.007 remain small and statistically weak.

Taken together, these findings support two interpretations. First, if water stress operates at the water-system level and is shared among consumers served by the same utility, it could explain a meaningful portion of the observed price response. Second, because water-service boundaries may

[10] Given data limitations—namely that the ACS provides electricity and water costs at the Census-tract level rather than for individual properties—we define treatment and control groups at the tract level. Treated tracts are those whose centroids lie within 5 km of a data center opening, while control tracts are those with centroids located 20–50 km from the same facility. Control variables include log population and log households, as well as tract demographics and socioeconomic characteristics: the shares of males, those under age 21, Blacks, Whites, college-educated (or above), employed, and those with household income below $25,000.

be sufficiently granular to capture other localized factors correlated with facility siting, we interpret the evidence cautiously. The attenuation is consistent with a water-stress channel, but it does not allow us to rule out additional localized disamenities that may also contribute to the baseline price declines.

5.4.3 Environmental Externality

Data center operations may depress nearby housing values in part through environmental disamenities. Plausible channels include intermittent on-site combustion from backup generator testing, emissions and drift associated with heat-rejection equipment, and meteorology-dependent propagation of mechanical noise from cooling infrastructure. A shared implication of these mechanisms is directional heterogeneity: for a given distance to the facility, expected exposure may be higher for locations that are more frequently downwind (Anderson (2020); Briggs (1975); Deryugina et al. (2019); Embleton (1996); International Organization for Standardization (2024); Ruiz et al. (2016)). Motivated by this, we implement a wind-direction diagnostic: if wind-transported externalities are a quantitatively important component of the capitalization effect, the post-opening price response should be more negative for parcels that are more often downwind of the facility.

To probe this possibility, we construct a parcel-level wind-propensity measure equal to the number of calendar months $(0 - 12)$ during which a parcel lies within a 60-degree downwind sector of its nearest data center, using long-run monthly-mean 10-meter wind vectors from the NARR. We then estimate a triple-difference specification in which the standard treated-by-post term is interacted with the wind propensity. The merit of this measure is that it provides a transparent, pre-determined proxy for expected directional exposure that is plausibly orthogonal to short-run housing market shocks. The limitation is equally clear: wind propensity does not measure any specific externality. It can proxy for wind-transported environmental elements (e.g., dispersion of airborne pollution, transmission of noise, drift and heat pathways, etc). At the same time, it is not informative about channels that do not travel with the wind (e.g., visual intrusion, traffic congestion). Accordingly, this test should be interpreted as reduced-form evidence on whether environmental externalities are a quantitatively relevant mechanism behind the price response.

Across specifications, the baseline capitalization effect remains similar in magnitude and economically meaningful: parcels within 5 km experience a statistically significant post-opening decline in sale prices of about $6.4 - 7.3\%$ (Table 6). However, the wind-based estimand does not reveal a clear directional pattern. Interpreted as the incremental post-opening effect per additional downwind month, the implied difference between parcels that are downwind in all twelve months versus never downwind is a 2.4% reduction in transaction prices. But the estimate is too imprecise to rule out either a negligible effect or a meaningfully larger one. As a placebo, we repeat the analysis using the upwind propensity, defined analogously as the number of calendar months a parcel lies in the opposite 60-degree sector. The corresponding estimate is approximately zero and also statistically insignificant.

Taken together, these results indicate that while prices decline significantly in the treated area after data center operation begins, we do not detect robust additional capitalization differentials along the prevailing downwind direction in this reduced-form exercise. This pattern can be explained by several possibilities: the relevant disamenity bundle may not be strongly wind-transported at the spatial and temporal resolution available here; directional effects may be attenuated by measurement error in wind propensity; or environmental channels may operate through pathways that are weakly directional. We, therefore, view the wind analysis as a suggestive diagnostic rather than a decisive test, and leave sharper attribution of mechanisms to future work that links facility operations to direct environmental measurements.

5.4.4 Alternative Mechanisms

Big Tech Effect. All hyperscale facilities in our estimation sample are owned and operated by big technology firms, including Amazon, Apple, Microsoft, Google, and Meta. To assess whether the estimated negative price effects reflect hyperscale characteristics (i.e., facility scale and intensity) or big-tech operator identity, Table 7 interacts a big-tech indicator with the DiD term using the full sample of data centers. Column (1) shows no detectable price response for non–big-tech-operated facilities, while the negative effect is concentrated among big-tech-operated data centers.

Because all hyperscale sites in our estimation sample are operated by big tech, the specification in Column (1) alone cannot disentangle big operator effects from scale effects. Column (2) therefore excludes hyperscale facilities and re-estimates the model on the remaining (non-hyperscale) data centers. Column (3) further excludes facilities with high utility power capacity, a proxy for larger-scale operations. The big tech interaction becomes statistically indistinguishable from zero, suggesting that operator identity per se is not the primary driver; instead, the negative price impacts appear to be associated with facility scale and hence the associated local disamenities.

Visual disamenity. If the facility's physical appearance were the dominant channel, housing prices should adjust immediately upon construction completion, when the structure becomes observable and salient to market participants. Instead, the price decline occurs only two years after completion. This timing pattern is difficult to reconcile with a purely aesthetic mechanism and is more consistent with disamenities that arise during operations or with gradual updating of beliefs as households learn about the facility's local externalities.

Infrastructure quality. A second possibility is that hyperscale data center development induces a reallocation of local infrastructure and public-service investment, with potential implications for broadly shared amenities (e.g., schools, hospitals, or other public services). We cannot fully dismiss such reallocations. However, this mechanism is unlikely to be the primary driver of our estimates. Infrastructure and service-provision channels typically operate at the county level and are therefore expected to generate relatively diffuse price effects within a county, rather than steep spatial attenuation. In contrast, our specification absorbs county-by-year shocks via county-by-year fixed effects, and the estimated impacts display a pronounced within-county distance gradient, with substantially larger declines for properties nearer to hyperscale data centers. This pattern suggests that, to the extent they occur, countywide investment shifts do not account for the localized effects we document.

5.5 Back-of-Envelope Calculation

We interpret the estimated housing-price response as the present value of households' revealed willingness to pay to avoid the localized externality bundle generated by hyperscale operations. Using the preferred pooled estimate, house prices within 5 km decline by 6.8% following commissioning. This magnitude is comparable to capitalization effects documented for other prominent locally undesirable facilities. For instance, Davis (2011) documented a $3-7\%$ reduction in housing values within 2 miles of newly opened fossil-fuel power plants. Our findings suggest that hyperscale data centers, despite not directly emitting air pollution like power plants, may generate other local disamenities that the housing market capitalizes in a similar order of magnitude. Using an average sale price of $\$350,000$, this effect corresponds to an implied loss of approximately $\$23,800$ per affected property. Scaling this per-property loss by an estimated 150,000 affected properties within 5 km of hyperscale facilities implies an aggregate capitalization loss of roughly \$3.6 billion in our sample.

6 Conclusion

This paper provides the first national-level causal evidence on the local housing market impacts of hyperscale data centers. Using a stacked difference-in-differences design across a comprehensive sample of data center operations, we find that proximity to hyperscale data centers induces economically meaningful declines in residential property values. In our preferred specification, housing prices within 5 km of a new hyperscale data centers decline by approximately 6.8%, with the effect concentrated within the first few kilometers. Event-time estimates show no differential pre-trends and a post-opening decline that strengthens over the subsequent years. This capitalization effect is stable across specifications and is driven primarily by the largest facilities operated by major technology firms.

We present complementary evidence to interpret the capitalization effect and assess potential channels. Consistent with hyperscale facilities being capital-intensive and labor-light, we find little evidence of sustained increases in local employment or wages following commissioning. In contrast, we document increases in residential electricity costs in host areas after commissioning, consistent with load-driven grid investments and/or procurement costs being partially passed through to households. We also examine whether the estimated housing-price effect is sensitive to water-system exposure by leveraging water-service boundary data. Absorbing water-service-boundary-by-time variation attenuates the estimated capitalization effect, which is consistent with a component of the effect operating through water-system-level conditions, although these specifications may also remove variation correlated with siting and local amenities. Finally, a wind-direction diagnostic yields negative but statistically imprecise differences between more- versus less-frequently downwind parcels, suggesting that wind-transported channels are unlikely to account for a large share of the baseline effect in this setting.

Our findings highlight a distinct feature of the modern digital economy—the spatial decoupling of benefits and costs. Because the estimated impacts decay sharply with distance, the physical footprint of hyperscale generates salient, hyper-localized welfare losses, while fiscal benefits from development deals accrue at the jurisdictional level that negotiates incentives. This wedge implies that evaluating subsidies solely in terms of aggregate investment or tax-base arguments can be misleading. Policies that explicitly account for incidence—including siting buffers from residential areas, enforceable nuisance standards (e.g., noise), and transparent rules for allocating incremental infrastructure costs—are likely to be more efficient and more equitable than uniform abatements that ignore localized burdens.

A stylized welfare calculation illustrates the stakes. Under the standard hedonic interpretation, the change in nearby house prices reflects households' willingness to pay to avoid the bundle of localized costs associated with hyperscale commissioning, net of any local benefits valued by residents (Bishop et al. (2020); Sheppard (1999)). Our estimates suggest that, absent significant fiscal transfers that are utilized to directly compensate affected homeowners, the net local welfare impact on nearby residents is negative. In particular, the drop in housing prices across the 150,000 affected properties observed in our data implies an aggregate welfare loss of approximately $3.6 billion. Consequently, local opposition to data center siting—often characterized as "NIMBYism"—may reflect a rational capitalization of net welfare losses rather than a rejection of technology itself.

Lastly, our analysis opens new avenues for future research. First, our donut-shaped design identifies localized effects net of any broader regional impacts shared by both treated and control areas; it does not speak to general-equilibrium benefits of digital infrastructure that accrue outside the near ring. Second, future research might fruitfully examine the precise elasticity of utility cost pass-through. Third, the externality bundle plausibly varies with grid conditions, cooling technology, climate, water constraints, and regulatory regimes. Linking openings to direct measures of noise, water withdrawals, grid upgrades, and local emissions would sharpen attribution of mechanisms and improve welfare accounting. As digital infrastructure expands, measuring and pricing its localized costs will be central to designing development strategies that internalize incidence rather than obscuring it. Our analysis establishes a clear baseline: under current technological and policy conditions, the cloud infrastructure casts a tangible shadow on local housing markets.

References

Abis, Simona, and Laura Veldkamp. 2024. "The changing economics of knowledge production." *Review of Financial Studies* 37 (1): 89–118.

Acemoglu, Daron, David Autor, Jonathon Hazell, and Pascual Restrepo. 2022. "Artificial intelligence and jobs: Evidence from online vacancies." *Journal of Labor Economics* 40 (S1): S293–S340.

Allcott, Hunt, and Daniel Keniston. 2018. "Dutch disease or agglomeration? The local economic effects of natural resource booms in modern America." *The Review of Economic Studies* 85 (2): 695–731.

Anderson, Michael L. 2020. "As the wind blows: The effects of long-term exposure to air pollution on mortality." *Journal of the European Economic Association* 18 (4): 1886–1927.

Babina, Tania, Anastassia Fedyk, Alex He, and James Hodson. 2024. "Artificial intelligence, firm growth, and product innovation." *Journal of Financial Economics* 151 103745.

Babina, Tania, Anastassia Fedyk, Alex X He, and James Hodson. 2023. *Firm investments in artificial intelligence technologies and changes in workforce composition.* Volume 31325. National Bureau of Economic Research.

Bartik, Alexander W, Janet Currie, Michael Greenstone, and Christopher R Knittel. 2019. "The local economic and welfare consequences of hydraulic fracturing." *American Economic Journal: Applied Economics* 11 (4): 105–155.

Basner, Mathias, Wolfgang Babisch, Adrian Davis, Mark Brink, Charlotte Clark, Sabine Janssen, and Stephen Stansfeld. 2014. "Auditory and non-auditory effects of noise on health." *The lancet* 383 (9925): 1325–1332.

Bishop, Kelly C, Nicolai V Kuminoff, H Spencer Banzhaf, Kevin J Boyle, Kathrine Von Gravenitz, Jaren C Pope, V Kerry Smith, and Christopher D Timmins. 2020. "Best practices for using hedonic property value models to measure willingness to pay for environmental quality." *Review of Environmental Economics and Policy.*

Blunt, Katherine, and Jennifer Hiller. 2026. "America's Biggest Power Grid Operator Has an AI Problem—Too Many Data Centers." January, https://www.wsj.com/business/energy-oil/power-grid-ai-data-centers-1235f296, The Wall Street Journal, January 13, 2026. Accessed 2026-01-14.

Borusyak, Kirill, Xavier Jaravel, and Jann Spiess. 2024. "Revisiting event-study designs: robust and efficient estimation." *Review of Economic Studies* 91 (6): 3253–3285.

Briggs, Gary A. 1975. "Plume rise predictions." In *Lectures on air pollution and environmental impact analyses*, 59–111, Springer.

Broner, Norman, and HG Leventhall. 1983. "Low frequency noise annoyance assessment by low frequency noise rating (LFNR) curves." *Journal of Low Frequency Noise, Vibration and Active Control* 2 (1): 20–28.

Bui, Linda TM, and Christopher J Mayer. 2003. "Regulation and capitalization of environmental amenities: evidence from the toxic release inventory in Massachusetts." *Review of Economics and statistics* 85 (3): 693–708.

Callaway, Brantly, and Pedro HC Sant'Anna. 2021. "Difference-in-differences with multiple time periods." *Journal of econometrics* 225 (2): 200–230.

Cao, Sean, Wei Jiang, Junbo Wang, and Baozhong Yang. 2024. "From man vs. machine to man+ machine: The art and AI of stock analyses." *Journal of Financial Economics* 160 103910.

Carlson, JP, and AM Teplitzky. 1974. "Environmental noise impact of natural-draft hyperbolic cooling towers." *The Journal of the Acoustical Society of America* 55 (S1): S36–S36.

Cary, Peter. 2023. "Amazon tones down its data center noise after residents sound the alarm." *Prince William Times*, October, https://www.princewilliamtimes.com/news/amazon-tones-down-its-data-center-noise-after-residents-sound-the-alarm/article_b81154ea-5178-59cb-ac3b-82228a2a52cf.html, Originally published via the Piedmont Journalism Foundation; accessed: 2025-12-09.

Chen, Mark A, Qinxi Wu, and Baozhong Yang. 2019. "How valuable is FinTech innovation?" *Review of financial studies* 32 (5): 2062–2106.

Cockburn, Iain M, Rebecca Henderson, and Scott Stern. 2018. "The impact of artificial intelligence on innovation: An exploratory analysis." In *The economics of artificial intelligence: An agenda*, 115–146, University of Chicago Press.

Currie, Janet, Lucas Davis, Michael Greenstone, and Reed Walker. 2015. "Environmental health risks and housing values: evidence from 1,600 toxic plant openings and closings." *American Economic Review* 105 (2): 678–709.

Davis, Lucas W. 2011. "The effect of power plants on local housing values and rents." *Review of Economics and Statistics* 93 (4): 1391–1402.

De Chaisemartin, Clément, and Xavier d'Haultfoeuille. 2024. "Difference-in-differences estimators of intertemporal treatment effects." *Review of Economics and Statistics* 1–45.

Deryugina, Tatyana, Garth Heutel, Nolan H Miller, David Molitor, and Julian Reif. 2019. "The mortality and medical costs of air pollution: Evidence from changes in wind direction." *American Economic Review* 109 (12): 4178–4219.

Dröes, Martijn I, and Hans RA Koster. 2016. "Renewable energy and negative externalities: The effect of wind turbines on house prices." *Journal of Urban Economics* 96 121–141.

D'Acunto, Francesco, Nagpurnanand Prabhala, and Alberto G Rossi. 2019. "The promises and pitfalls of robo-advising." *Review of Financial Studies* 32 (5): 1983–2020.

Ellis, RM. 1971. "Cooling tower noise generation and radiation." *Journal of Sound and Vibration* 14 (2): 171–182.

Embleton, Tony FW. 1996. "Tutorial on sound propagation outdoors." *The Journal of the Acoustical Society of America* 100 (1): 31–48.

Feher, Adam, Emilia Garcia-Appendini, and Roxana Mihet. 2025. "Is AI Trained on Public Money? Evidence from US Data Centers." *Evidence from US Data Centers (September 05, 2025). Swiss Finance Institute Research Paper* (25-73): .

Fraenkel, Rebecca, Josh Graff Zivin, and Sam Krumholz. 2024. "The coal transition and its implications for health and housing values." *Land Economics* 100 (1): 51–65.

Fuster, Andreas, Paul Goldsmith-Pinkham, Tarun Ramadorai, and Ansgar Walther. 2022. "Predictably unequal? The effects of machine learning on credit markets." *The Journal of Finance* 77 (1): 5–47.

Gopalakrishnan, Sathya, and H Allen Klaiber. 2014. "Is the shale energy boom a bust for nearby residents? Evidence from housing values in Pennsylvania." *American Journal of Agricultural Economics* 96 (1): 43–66.

Greenstone, Michael, and Justin Gallagher. 2008. "Does hazardous waste matter? Evidence from the housing market and the superfund program." *The Quarterly Journal of Economics* 123 (3): 951–1003.

Grennan, Jillian, and Roni Michaely. 2020. "Artificial intelligence and high-skilled work: Evidence from analysts." *Swiss Finance Institute Research Paper* (20-84): .

Hamilton, Stanley W, and Gregory M Schwann. 1995. "Do high voltage electric transmission lines affect property value?" *Land economics* 436–444.

Heintzelman, Martin D, and Carrie M Tuttle. 2012. "Values in the wind: A hedonic analysis of wind power facilities." *Land Economics* 88 (3): 571–588.

Hirvonen, Johannes, Aapo Stenhammar, and Joonas Tuhkuri. 2022. *New evidence on the effect of technology on employment and skill demand*. Taloustieto Oy.

Holland, Stephen P, and Erin T Mansur. 2008. "Is real-time pricing green? The environmental impacts of electricity demand variance." *The Review of Economics and Statistics* 90 (3): 550–561.

Hu, Chenyang, Zhenshan Chen, Pengfei Liu, Wei Zhang, Xi He, and Darrell Bosch. 2025. "Impact of large-scale solar on property values in the United States: Diverse effects and causal mechanisms." *Proceedings of the National Academy of Sciences* 122 (24): e2418414122.

International Organization for Standardization. 2024. "Acoustics — Attenuation of sound during propagation outdoors — Part 2: Engineering method for the prediction of sound pressure levels outdoors." International Standard ISO 9613-2:2024, International Organization for Standardization, Geneva, Switzerland, https://www.iso.org/standard/74047.html, Edition 2. Published 2024-01-12. Accessed 2026-01-14.

Jansen, Mark, Hieu Nguyen, and Amin Shams. 2020. *Human vs. machine: Underwriting decisions in finance*. Ohio State University, Fisher College of Business, Charles A. Dice Center ....

Kline, Patrick, and Enrico Moretti. 2014. "People, places, and public policy: Some simple welfare economics of local economic development programs." *Annu. Rev. Econ.* 6 (1): 629–662.

Miedema, Henk ME, and Henk Vos. 2003. "Noise sensitivity and reactions to noise and other environmental conditions." *The journal of the acoustical society of America* 113 (3): 1492–1504.

Moretti, Enrico. 2010. "Local multipliers." *American Economic Review* 100 (2): 373–377.

Muehlenbachs, Lucija, Elisheba Spiller, and Christopher Timmins. 2015. "The housing market impacts of shale gas development." *American Economic Review* 105 (12): 3633–3659.

Muzet, Alain. 2007. "Environmental noise, sleep and health." *Sleep medicine reviews* 11 (2): 135–142.

Mytton, David. 2021. "Data centre water consumption." *npj Clean Water* 4 (1): 11.

Offutt, Martin C., and Ling Zhu. 2025. "Data Centers and Their Energy Consumption: Frequently Asked Questions."Technical Report R48646, Congressional Research Service, Washington, DC.

Rock, Daniel. 2019. "Engineering value: The returns to technological talent and investments in artificial intelligence." *Working Paper*.

Ruiz, J, CG Cutillas, AS Kaiser, M Ballesta, B Zamora, and M Lucas. 2016. "Experimental study of drift deposition from mechanical draft cooling towers in urban environments." *Energy and Buildings* 125 181–195.

Shehabi, Arman, Sarah J. Smith, Alex Hubbard et al. 2024. "2024 United States Data Center Energy Usage Report."Technical Report LBNL-2001637, Lawrence Berkeley National Laboratory, Berkeley, California, https://eta.lbl.gov/publications/2024-lbnl-data-center-energy-usage-report, Accessed: 2025-12-09.

Sheppard, Stephen. 1999. "Hedonic analysis of housing markets." *Handbook of regional and urban economics* 3 1595–1635.

Siddik, Md Abu Bakar, Arman Shehabi, and Landon Marston. 2021. "The environmental footprint of data centers in the United States." *Environmental Research Letters* 16 (6): 064017.

Sun, Liyang, and Sarah Abraham. 2021. "Estimating dynamic treatment effects in event studies with heterogeneous treatment effects." *Journal of econometrics* 225 (2): 175–199.

Wade, Cameron, Mike Blackhurst, Joe DeCarolis, Anderson de Queiroz, Jeremiah Johnson, and Paulina Jaramillo. 2025. "Electricity Grid Impacts of Rising Demand from Data Centers and Cryptocurrency Mining Operations." https://energy.cmu.edu/_files/documents/electricity-grid-impacts-of-rising-demand-from-data-centers-and-cryptocurrency-mining-operat pdf.

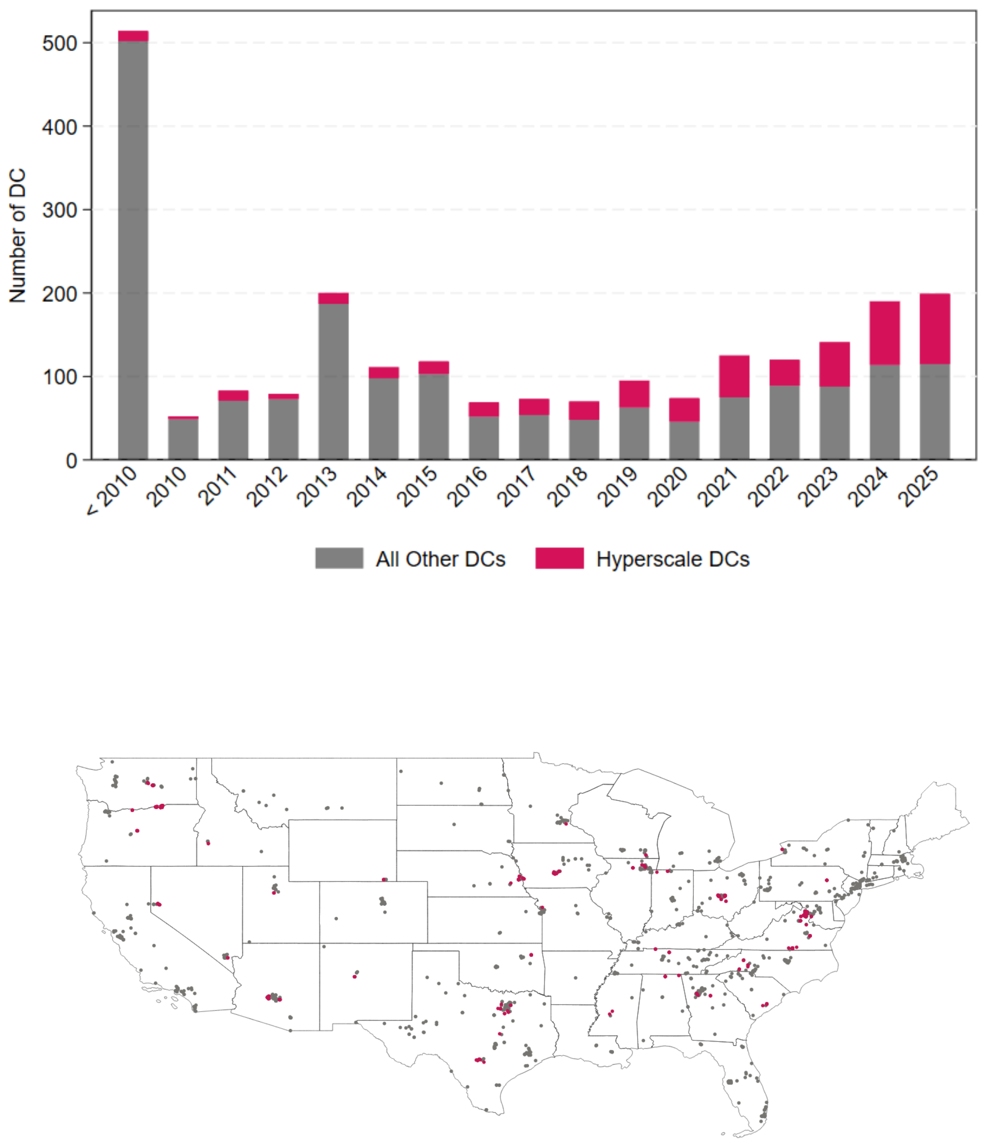
Zheng, Siqi, Weizeng Sun, Jianfeng Wu, and Matthew E Kahn. 2017. "The birth of edge cities in China: Measuring the effects of industrial parks policy." *Journal of Urban Economics* 100 80–103.

Zhouo, Yang, Ming Gao, Suoying He, Yuetao Shi, and Fengzhong Sun. 2020. "Case study: Theoretical calculation model and variable condition analysis research on the water-splashing noise for natural draft wet cooling towers." *Noise Control Engineering Journal* 68 (2): 137–145.

Zou, Eric. 2020. "Wind turbine syndrome: The impact of wind farms on suicide." *Working Paper*.

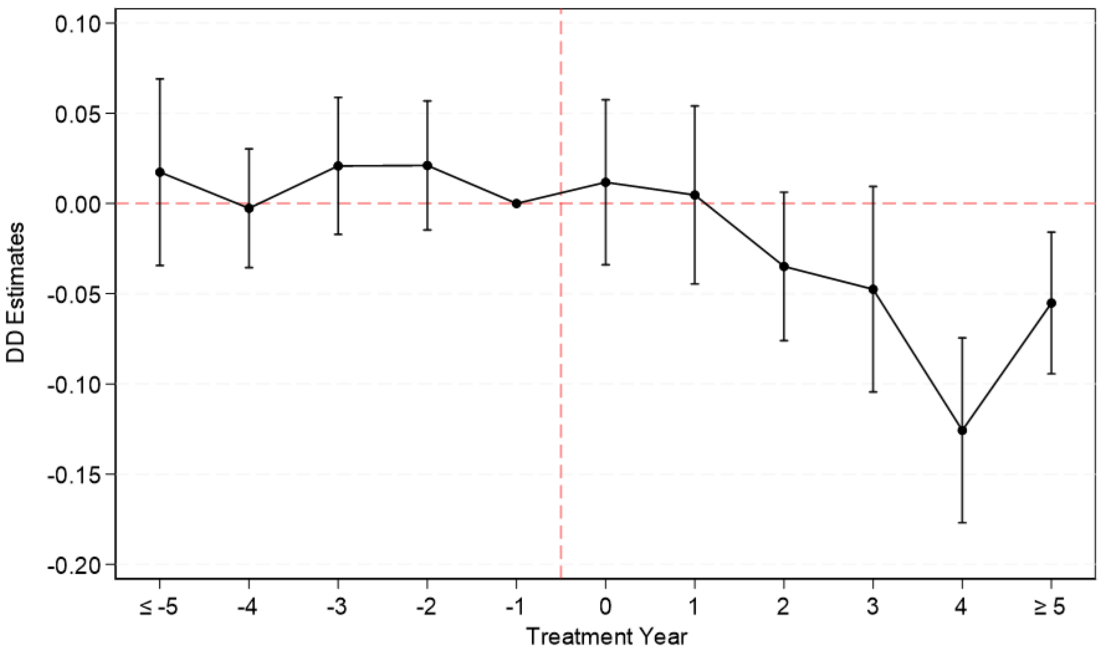
7 Figures and Tables

Figure 1: Data Center Development over Time and across the United States



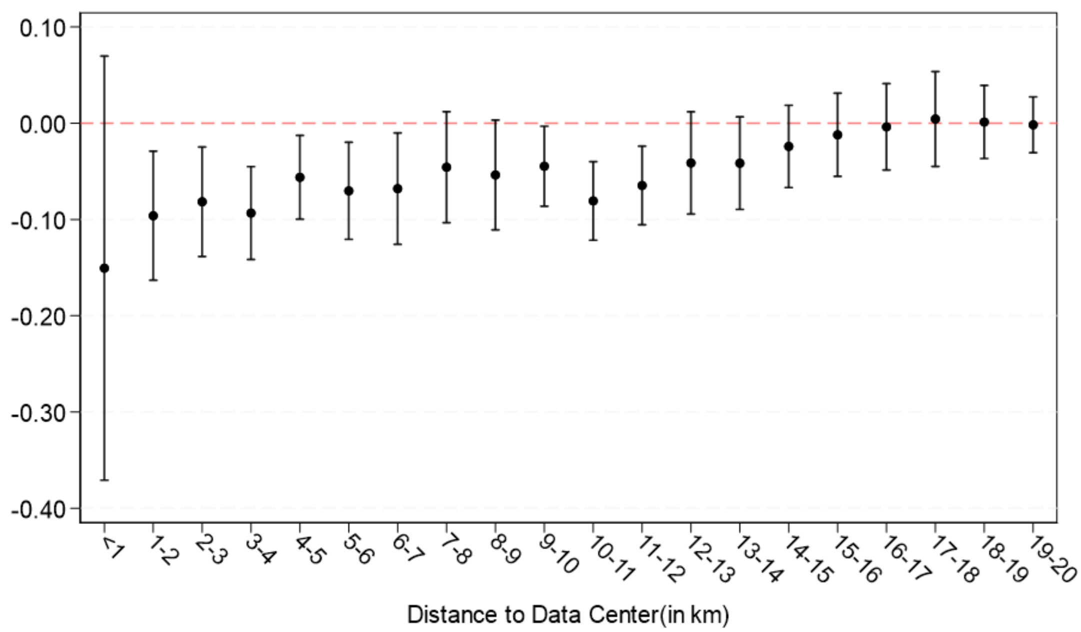
Notes: The figure plots the number of data centers by construction-completion year (top panel) and their geographic distribution across the contiguous United States (bottom panel). Red bars in the top panel and red dots in the bottom panel denote hyperscale data centers, while the gray counterparts represent all other facilities, including retail, wholesale, and other types.

Figure 2: Event Study of Log Housing Prices Around Hyperscale Data Center Openings



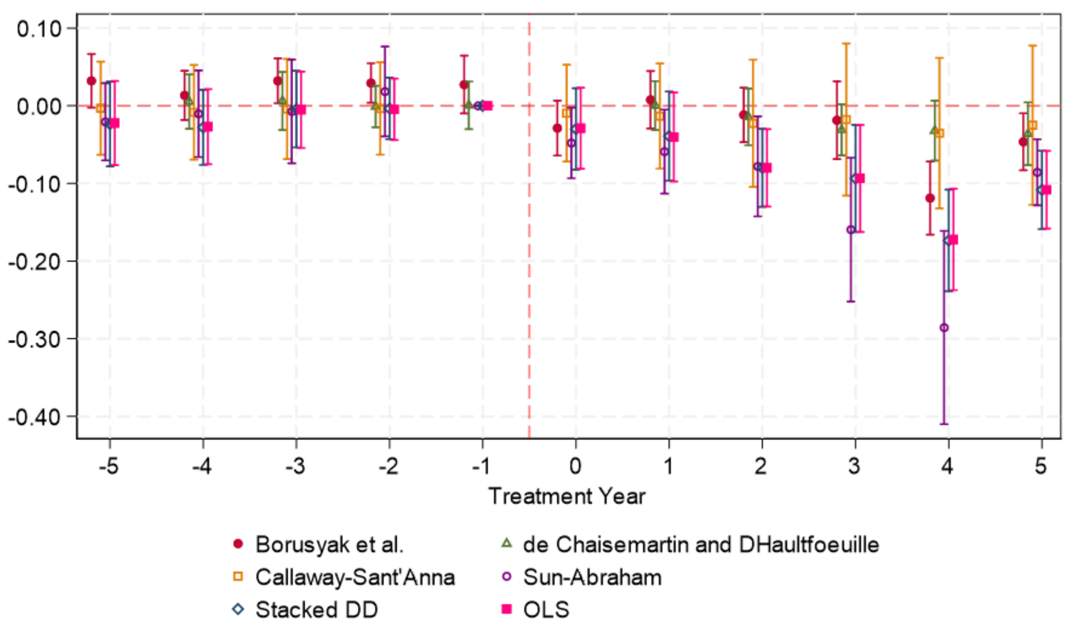
Notes: This figure plots the estimated event-time coefficients for log sale price from the baseline difference-in-differences model specified in Equation 2. The horizontal axis is years relative to the data center opening (year 0 = opening year). The point at -1 (the omitted category) is normalized to zero. Dots represent point estimates $\hat{\beta}_\ell$ and the lines show 95% confidence intervals.

Figure 3: Housing Price Impact by Distance from Hyperscale Data Center



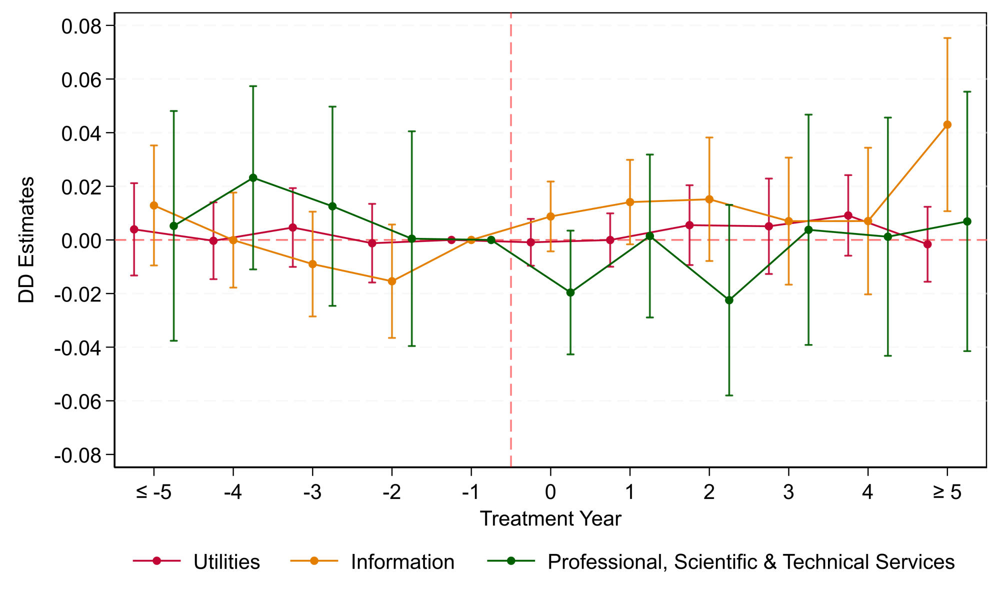
Notes: This figure shows estimated treatment effects on log housing price for different distance bands around the data center as specified in Equation 2. Dots represent point estimates $\hat{\beta}_\ell$ and the lines show 95% confidence intervals.

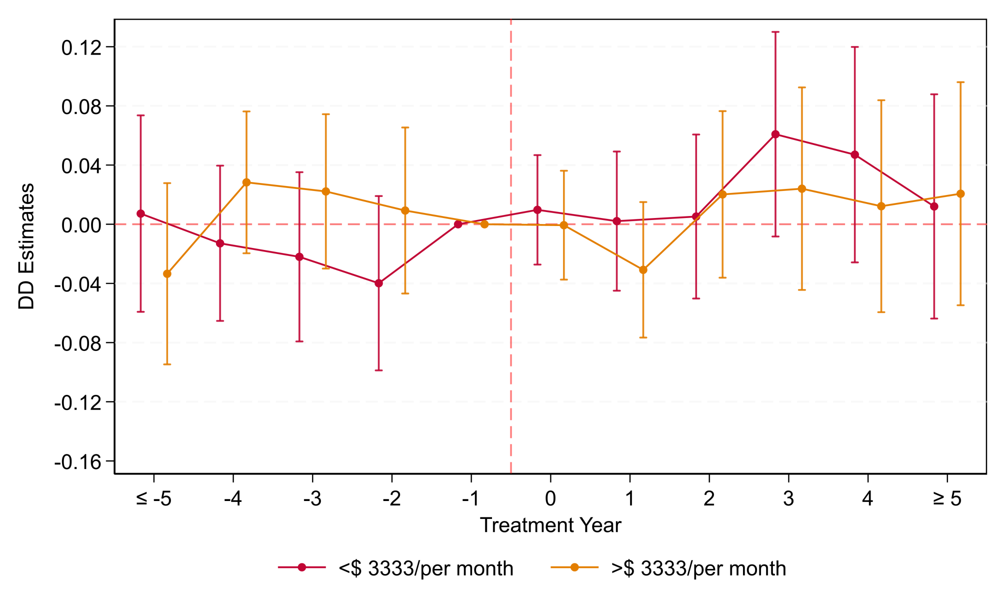
Figure 4: Staggered DiD Results of Log Housing Prices Around Hyperscale Data Center Openings



Notes: This figure plots the estimated event-time coefficients for log sale price from various staggered difference-in-differences models. The horizontal axis is years relative to the data center opening (year 0 = opening year). The point at -1 (the omitted category) is normalized to zero. Dots represent point estimates $\hat{\beta}_\ell$ and the lines show 95% confidence intervals.

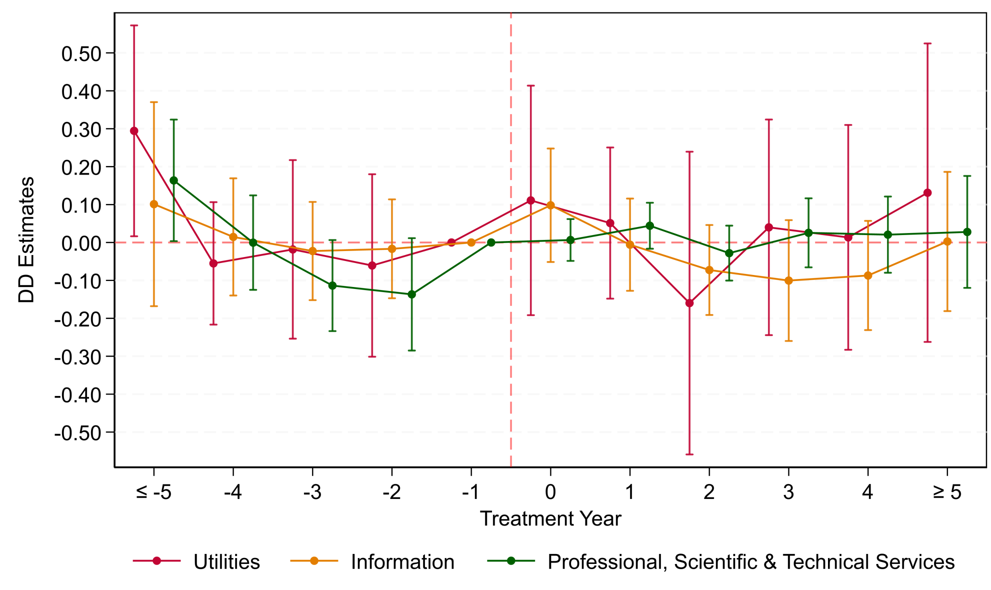
Figure 5: Labor Market Impact of Hyperscale Data Center

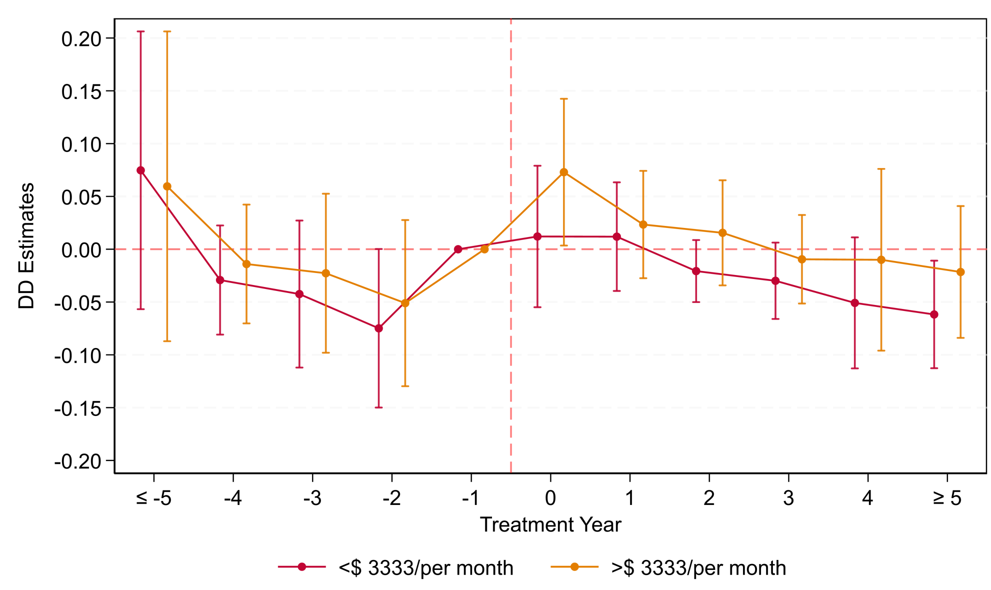




Notes: This figure plots event-time coefficients for labor-market outcomes in natural logarithms. The horizontal axis is years relative to the data center opening (year 0 = opening year). The point at -1 (the omitted category) is normalized to zero. Dots represent point estimates $\hat{\beta}_\ell$ and the lines show 95% confidence intervals.

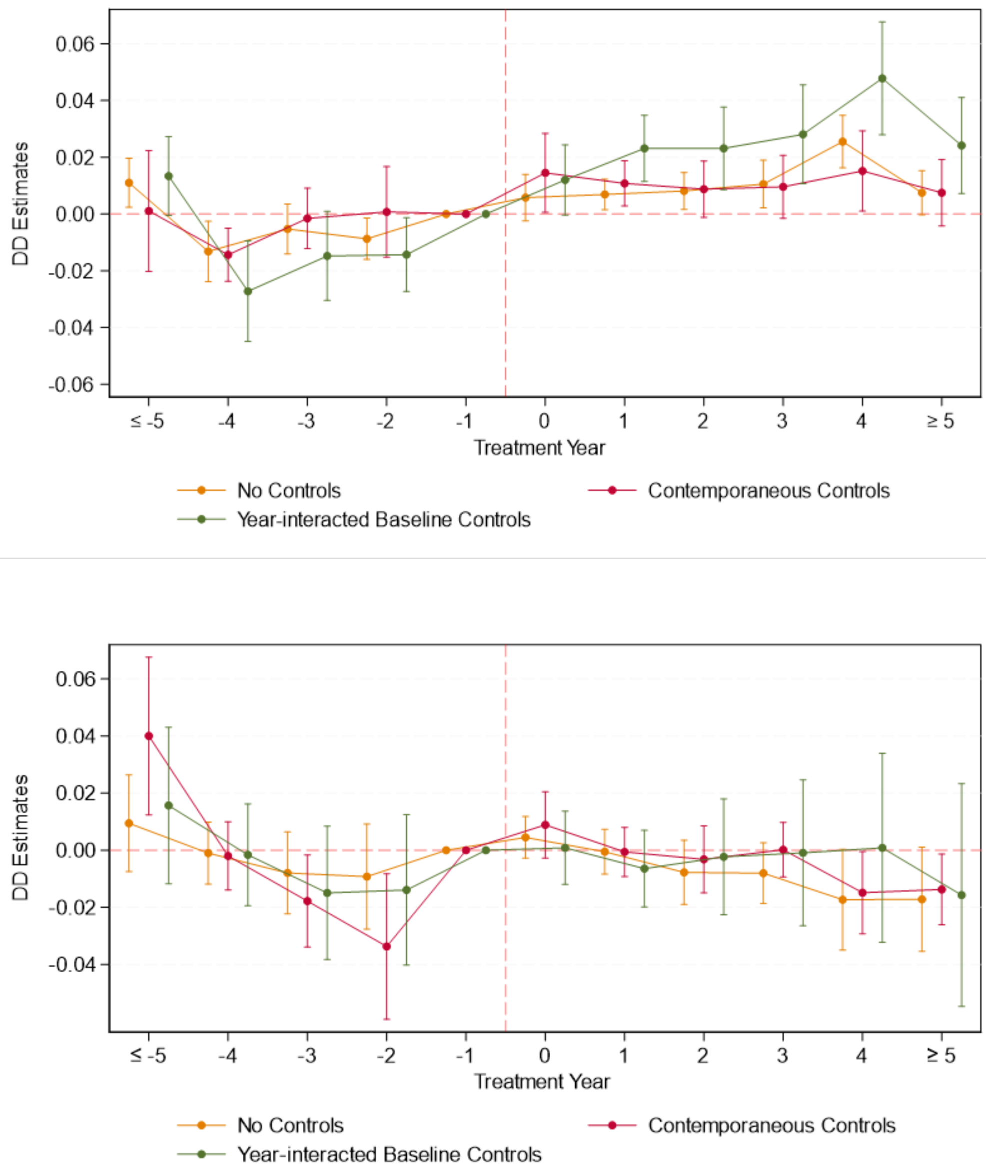
Figure 6: Labor Market Impact of Hyperscale Data Center—at County Level





Notes: This figure plots event-time coefficients for labor-market outcomes, aggregated at the county level, in natural logarithms. The horizontal axis is years relative to the data center opening (year 0 = opening year). The point at -1 (the omitted category) is normalized to zero. Dots represent point estimates $\hat{\beta}_\ell$ and the lines show 95% confidence intervals.

Figure 7: Utillity Cost Impact of Hyperscale Data Center



Notes: This figure plots the estimated event-time coefficients for electricity (top panel) and water (bottom panel) utility cost from the stacked difference-in-differences model. The horizontal axis is years relative to the data center opening (year 0 = opening year). The point at -1 (the omitted category) is normalized to zero. Dots represent point estimates $\hat{\beta}_\ell$ and the lines show 95% confidence intervals.

Table 1: Summary Statistics of Property Transaction Characteristics by Distance to Data Centers

	Mean Values of All Data Center		Mean Values of Hyperscale Data Center	
	≤ 5km	20 - 50km	≤ 5km	20 - 50km
Transaction Characteristics				
Sale price ($'000)	446.944	380.410	349.070	364.601
Distance to data center (km)	3.437	37.774	3.687	38.163
Local Buyer	0.810	0.768	0.841	0.791
Corporate Buyer	0.097	0.075	0.064	0.077
Local Seller	0.810	0.754	0.839	0.767
Corporate Seller	0.381	0.396	0.451	0.358
Cash Purchase	0.201	0.198	0.114	0.181
Resale	0.876	0.822	0.737	0.839
Short Sale	0.022	0.020	0.020	0.018
Foreclosure Sale	0.066	0.060	0.054	0.053
Property Characteristics				
Living Space ('000sqf)	2.025	2.419	2.056	2.049
No. Bathrooms	2.283	2.396	2.582	2.356
No. Bedrooms	3.165	3.205	3.234	3.154
No. Stories	1.474	1.411	1.476	1.407
No. Parking Spaces	2.614	2.516	2.133	2.016
Pool	0.073	0.082	0.061	0.059
Property Age (Years)	37.119	25.358	20.752	26.609
Neighborhood Demographics				
Total population ('000)	9.763	11.002	11.093	11.015
Total households ('000)	3.684	3.972	3.893	4.066
Total housing units ('000)	4.076	4.454	4.184	4.556
Black Share	0.147	0.078	0.111	0.107
Non-Hispanic White Share	0.696	0.829	0.727	0.790
Median Household Income ($'000)	64.126	67.954	72.779	66.331
Unemployment Share	0.032	0.028	0.027	0.026
Observations	4,580,195 (11%)	35,726,876 (88%)	221,986 (6%)	3,479,573 (94%)

Notes: The unit of analysis is at the property transaction level. Transactions are grouped by distance to the nearest data center site: properties within 5 km versus properties $20 - 50$ km away.

Table 2: Effects of Data Centers on House Prices – Stacked DiD

	Log Housing Sale Price					
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A: Effects of All Data Centers						
$\leq$ 5km $\times$ Post	0.002	-0.002	-0.008^{**}	-0.011^{***}	-0.020^{***}	0.001
	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)	(0.005)
Adj. R^2	0.503	0.620	0.550	0.649	0.652	0.675
Observations	40,307,071	40,306,824	40,307,071	40,306,824	40,297,430	37,338,946
Panel B: Effects of Hyperscale Data Centers						
$\leq$ 5km $\times$ Post	-0.041^{**}	-0.060^{***}	-0.030^{**}	-0.046^{***}	-0.049^{***}	-0.068^{***}
	(0.016)	(0.017)	(0.014)	(0.014)	(0.016)	(0.021)
Adj. R^2	0.478	0.607	0.521	0.640	0.662	0.643
Observations	3,701,559	3,701,481	3,701,559	3,701,481	3,699,640	3,661,868
Transaction Controls	No	No	Yes	Yes	Yes	Yes
Property Controls	No	Yes	No	Yes	Yes	Yes
DC $\times$ Distance FE	Yes	Yes	Yes	Yes	Yes	Yes
DC $\times$ Census Tract FE	Yes	Yes	Yes	Yes	Yes	Yes
DC $\times$ Time FE	Yes	Yes	Yes	Yes	No	No
DC $\times$ State $\times$ Time FE	No	No	No	No	Yes	No
DC $\times$ County $\times$ Time FE	No	No	No	No	No	Yes

Notes: Each column reports estimates from the stacked difference-in-differences specification around data center openings, comparing properties within 5 km to properties in the control ring. Parentheses report standard errors.
***Significant at the 1 percent level.
**Significant at the 5 percent level.
*Significant at the 10 percent level.

Table 3: Alternative Control and Treatment Groups

	Log Housing Sale Price					
	(1)	**(2)**	**(3)**	**(4)**	**(5)**	**(6)**
Treated × Post	-0.034^{***}	-0.056^{**}	-0.053^{***}	-0.089^{***}	-0.047^{***}	-0.016^{*}
	(0.012)	(0.013)	(0.016)	(0.022)	(0.017)	(0.008)
Transaction Controls	Yes	Yes	Yes	Yes	Yes	Yes
Property Controls	Yes	Yes	Yes	Yes	Yes	Yes
DC × Census Tract FE	Yes	Yes	Yes	Yes	Yes	Yes
DC × County × Time FE	Yes	Yes	Yes	Yes		
DC × Distance-to-DC FE	Yes	Yes	Yes	Yes	Yes	No
Distance-to-DC FE	No	No	No	No	Yes	No
Adj. R^2	0.686	0.667	0.676	0.667	0.702	0.666
Observations	27,834,778	6,013,718	1,568,211	3,119,525	25,022,496	568,382
Treatment Group	≤5 km	≤5 km	≤5 km	≤5 km	≤5 km	CTs with DC
Control Group	5–50 km	15–50 km	15–30 km	30–50km	Later-treated ≤5 km	Neigh. CTs w/o DC

Notes: Each column reports a stacked difference-in-differences estimate of the effect of data center openings on log housing sale prices under an alternative definition of treatment and/or the comparison group. The bottom panel lists the treatment and control groups used in each specification. Parentheses report standard errors.

***Significant at the 1 percent level.

**Significant at the 5 percent level.

*Significant at the 10 percent level.

Table 4: Alternative Definitions of Hyperscale Data Centers

	Log Housing Sale Price					
	(1)	**(2)**	**(3)**	**(4)**	**(5)**	**(6)**
≤5km × Post	−0.135***	−0.070***	−0.016	−0.004	0.029*	0.001
	(0.029)	(0.022)	(0.019)	(0.014)	(0.015)	(0.005)
Transaction Controls	Yes	Yes	Yes	Yes	Yes	Yes
Property Controls	Yes	Yes	Yes	Yes	Yes	Yes
DC × Distance-to-DC FE	Yes	Yes	Yes	Yes	Yes	Yes
DC × Census Tract FE	Yes	Yes	Yes	Yes	Yes	Yes
DC × County × Time FE	Yes	Yes	Yes	Yes	Yes	Yes
Adj. R^2	0.665	0.657	0.665	0.670	0.678	0.675
Observations	823,233	995,139	2,363,553	5,250,730	8,912,229	37,338,946
Total Utility Power (MW)	≥100	≥80	≥60	≥40	≥20	All

Notes: Each column reports a stacked difference-in-differences estimate of the effect of data center openings on log housing sale prices, redefining the definition of "hyperscale" sites using alternative thresholds of total utility power capacity. Parentheses report standard errors.

***Significant at the 1 percent level.

**Significant at the 5 percent level.

*Significant at the 10 percent level.

Table 5: Water Stress and Housing Price Effects

	Log Housing Sale Price			
	(1)	(2)	(3)	(4)
$\leq$ 5km $\times$ Post	-0.014	-0.009	-0.042*	-0.007
	(0.020)	(0.020)	(0.016)	(0.019)
Transaction Controls	Yes	Yes	Yes	Yes
Property Controls	Yes	Yes	Yes	Yes
DC $\times$ Distance FE	Yes	Yes	Yes	Yes
DC $\times$ Census Tract FE	Yes	Yes	Yes	Yes
DC $\times$ Service Boundary $\times$ Time FE	Yes	No	Yes	No
DC $\times$ County $\times$ Service Boundary $\times$ Time FE	No	Yes	No	Yes
Adj. R^2	0.703	0.705	0.672	0.686
Observations	2,241,747	2,231,815	3,551,924	3,491,918
Sample	Transactions Matched to TEMM		All	

Notes: Each column reports estimates from the stacked difference-in-differences specification around data center openings, comparing properties within 5 km to properties in the control ring. Parentheses report standard errors.
***Significant at the 1 percent level.
**Significant at the 5 percent level.
*Significant at the 10 percent level.

Table 6: Wind Direction and Housing Price Effects

	Log Housing Sale Price	
	(1)	(2)
$\leq$ 5km $\times$ Post	-0.064^{***}	-0.073^{***}
	(0.020)	(0.021)
$\leq$ 5km $\times$ Post $\times$ Downwind Months	-0.002	
	(0.003)	
$\leq$ 5km $\times$ Post $\times$ Upwind Months		0.000
		(0.004)
Transaction Controls	Yes	Yes
Property Controls	Yes	Yes
DC $\times$ Distance FE	Yes	Yes
DC $\times$ Census Tract FE	Yes	Yes
DC $\times$ County $\times$ Time FE	Yes	Yes
Adj. R^2	0.662	0.662
Observations	3,661,868	3,661,868
Sample	Hyperscale DC	Hyperscale DC

Notes: "Downwind Months" and "Upwind Months" indicate the number of calendar months the land parcels are in the downwind/upwind sector of the nearby data center. Parentheses report standard errors.
***Significant at the 1 percent level.
**Significant at the 5 percent level.
*Significant at the 10 percent level.

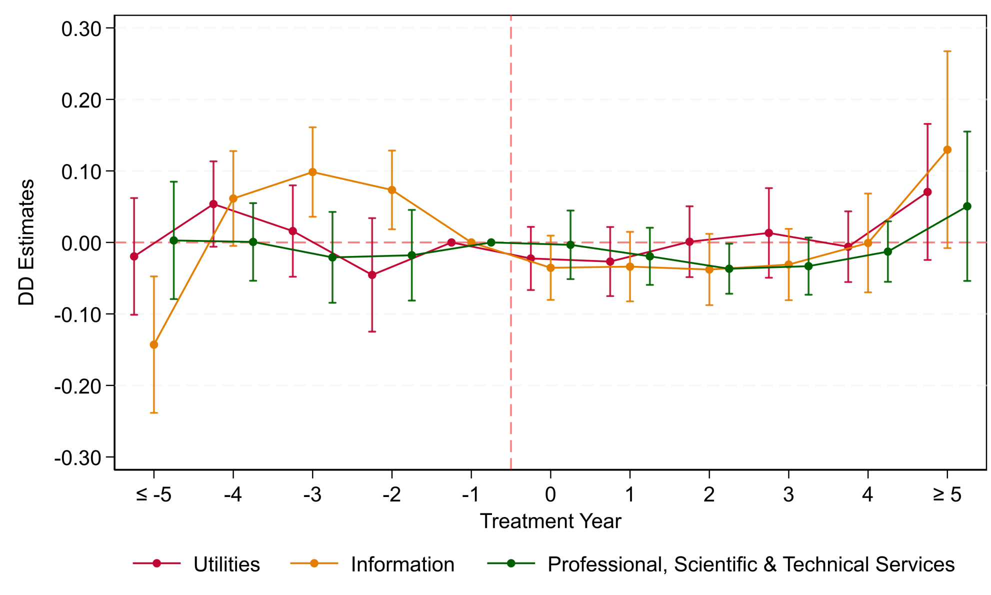
Table 7: Big Tech Ownership and Housing Price Effects

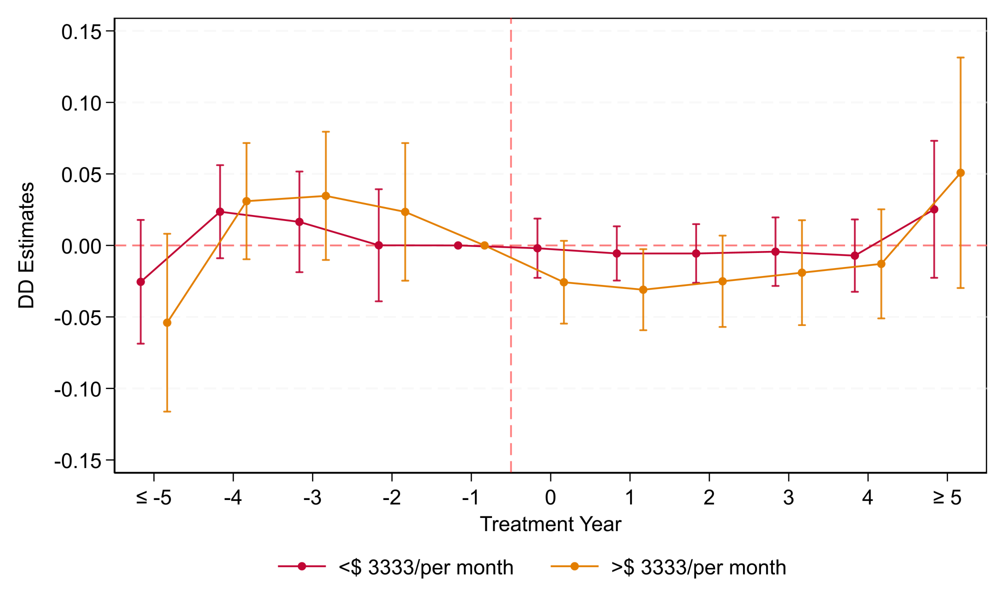
	Log Housing Sale Price		
	(1)	(2)	(3)
$\leq$ 5km $\times$ Post	0.003	0.003	0.003
	(0.005)	(0.005)	(0.005)
$\leq$ 5km $\times$ Post $\times$ Big Tech	-0.063^{***}	0.104	0.110
	(0.020)	(0.071)	(0.074)
Transaction Controls	Yes	Yes	Yes
Property Controls	Yes	Yes	Yes
DC $\times$ Distance FE	Yes	Yes	Yes
DC $\times$ Census Tract FE	Yes	Yes	Yes
DC $\times$ County $\times$ Time FE	Yes	Yes	Yes
Adj. R^2	0.675	0.678	0.678
Observations	37,338,946	33,616,358	31,434,552
Sample	All	w/o Hyperscale DC	w/o Hyperscale DC or DC with power < 80 MW

Notes: "Big Tech" indicates data centers operated by Amazon, Meta, Google, Apple, or Microsoft. Parentheses report standard errors.
***Significant at the 1 percent level.
**Significant at the 5 percent level.
*Significant at the 10 percent level.

A Appendix Figures and Tables

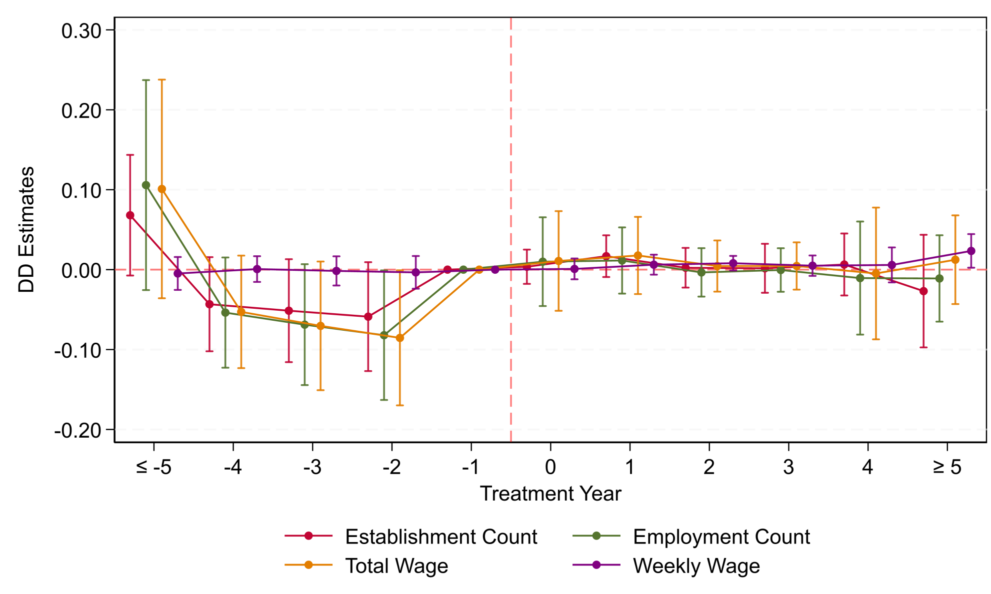
Figure A-1: Labor Market Impact of Hyperscale Data Center—at Commuting Zone Level

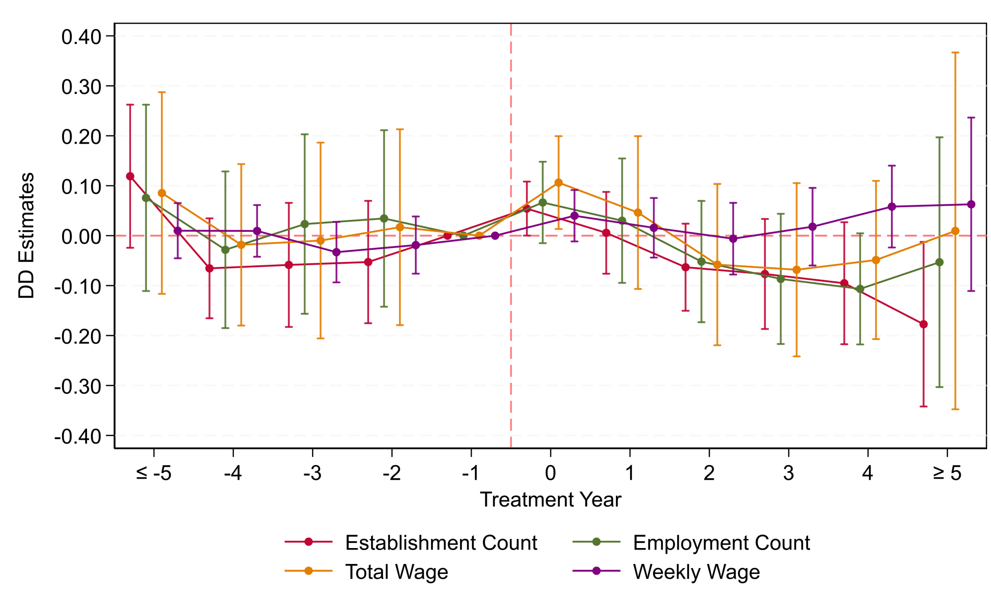




Notes: This figure plots event-time coefficients for labor-market outcomes, aggregated at the commuting zone level, in natural logarithms. The horizontal axis is years relative to the data center opening (year 0 = opening year). The point at -1 (the omitted category) is normalized to zero. Dots represent point estimates $\hat{\beta}_\ell$ and the lines show 95% confidence intervals.

Figure A-2: Labor Market Impact of Hyperscale Data Center—County-level QCEW





Notes: This figure plots event-time coefficients for labor-market outcomes, obtained from QCEW, in natural logarithms. The horizontal axis is years relative to the data center opening (year 0 = opening year). The point at -1 (the omitted category) is normalized to zero. Dots represent point estimates $\hat{\beta}_\ell$ and the lines show 95% confidence intervals.

Table A-1: Effects of Smaller-Scale Data Centers on House Prices

	Log Housing Sale Price				
	(1)	(2)	(3)	(4)	(5)
Treatment × Post	−0.017	0.007	0.002	−0.003	0.000
	(0.015)	(0.007)	(0.008)	(0.006)	(0.005)
Transaction Controls	Yes	Yes	Yes	Yes	Yes
Property Controls	Yes	Yes	Yes	Yes	Yes
DC × Distance FE	Yes	Yes	Yes	Yes	Yes
DC × Census Tract FE	Yes	Yes	Yes	Yes	Yes
DC × County × Time FE	Yes	Yes	Yes	Yes	Yes
Adj. R^2	0.673	0.674	0.674	0.676	0.677
Observations	29,558,420	29,888,662	30,244,787	30,587,720	31,068,396
Treatment Group	≤ 1km	1–2km	2–3km	3–4km	4–5km
Control Group			20–50km		

Notes: The sample includes non-hyperscale data centers. Each column reports a stacked difference-in-differences estimate using a treated ring indicated in the bottom panel and a common control ring of $20-50$ km. Parentheses report standard errors.
***Significant at the 1 percent level.
**Significant at the 5 percent level.
*Significant at the 10 percent level.