The Spatial Incidence of Hyperscale Data Centers[*]

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

April 18, 2026
First Draft: December 21, 2025

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 Please do not cite without permission.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

Abstract

We estimate the local incidence of U.S. data center openings using a stacked difference-in-differences design and nationwide housing transaction data. We find that hyperscale data centers—facilities built, owned and used by large cloud and AI operators—reduce nearby house prices by 6.8 percent, with effects highly localized and fading beyond 14 km. These effects increase with power capacity and attenuate with prior local exposure. By contrast, large non-hyperscale facilities show little evidence of capitalization on average, although negative effects emerge when they are sited in more residentially exposed environments. These results imply that the local incidence of digital infrastructure depends not only on scale, but also on siting context, local exposure, and salience.

JEL Codes: L86, Q51, R31

---

[*] We would like to thank Corbett Grainger, Daniel Phaneuf, Chris Timmins, Yu Qin, and seminar participants at the University of Wisconsin-Madison and National University of Singapore for helpful comments. All errors are ours.

[†] 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

State and local governments increasingly subsidize digital infrastructure in view of potential local economic gains. A leading example is the hyperscale data center: a large, power-intensive facility operated by major technology firms to support cloud computing and artificial intelligence. These projects often receive tax incentives, zoning accommodations, and other place-based support. Yet they differ from the large industrial plants that anchor the traditional development literature. Hyperscale facilities are highly capital-intensive, create little permanent local employment, and require substantial land, electricity, and water infrastructure. They may also impose localized costs through grid strain, water demand, noise, and other environmental externalities (e.g., Mytton, 2021; Shehabi et al., 2024; Siddik et al., 2021). This raises a central policy question: when localities attract data centers, do nearby residents share in the gains, or do they bear costs that are concentrated near the site?

This paper studies that question through housing-market capitalization. We provide national causal evidence on the localized incidence of data center openings, with a particular focus on hyperscale facilities, using housing prices to measure households' valuation of the net local effects of these investments. The effect is ex ante ambiguous. Data centers may expand the local tax base and attract complementary investment, but such benefits may accrue broadly across the jurisdiction while costs remain highly localized. We, therefore, ask whether proximity to data centers is capitalized as a net local amenity or a net local burden, and more broadly, whether the benefits and costs of digital infrastructure are spatially aligned.

To study these effects, we combine a proprietary dataset on nationwide data center locations, ownership, commissioning dates, and technical characteristics with nationwide residential transaction records and local economic data. The central empirical challenge is that data center siting is not random. We address this in three ways. First, for each opening, we compare housing transactions in a near ring (within 5 km) with transactions in an outer ring (20–50 km) around the same facility, within the same county and year-month. Second, to address the concern that locations near eventual data center sites may differ systematically even within counties, we compare treated near rings to the near rings of sites that open later. Third, we refine this approach by matching later-treated sites to current openings using XGBoost-predicted siting propensity scores estimated from county-and tract-level pre-treatment predictors of siting, including subsidies, permitting conditions, land availability, and local infrastructure.[1] This design compares current openings to observationally similar locations that were also at risk of treatment but had not yet opened. Across specifications, identification relies on the assumption that, absent the opening, housing prices near treated sites would have evolved similarly to either those farther from the same site or those near later-treated sites, conditional on a rich set of controls and fixed effects.

We find that hyperscale data center openings reduce nearby transaction prices by 6.8 percent. The estimate is stable across a range of specifications and alternative comparison groups. The effect is highly localized: it is largest closest to the site, reaching roughly 15% within 1 km, and attenuates with distance, becoming muted beyond about 14 km. Event-time estimates show no evidence of differential pre-trends and indicate that prices decline with a lag, turning negative about two years after opening and remaining below zero thereafter. By contrast, when we estimate the same design for *all* data centers, the effects are small and imprecisely estimated.

Additional heterogeneity evidence indicates that both operational capacity and local exposure shape the localized incidence of data center development. Within the hyperscale segment, capitalization effects are larger for more power-intensive sites. By contrast, large non-hyperscale facilities show little evidence of capitalization on average, even at high capacity thresholds. A key difference lies in siting context: hyperscale facilities are more often located in areas with greater residential land and less commercial, industrial, and government land, whereas non-hyperscale facilities are more commonly embedded in existing mixed-use activity centers.[2] Consistent with a role for local exposure and salience, the average null for large non-hyperscale facilities masks substantial heterogeneity: when such facilities are located in more residentially exposed environments, nearby house prices also decline sharply. We find similar evidence along other salience margins. First, we show that local Google Search intensity rises much more around hyperscale openings than around other facility types. Second, price effects attenuate with prior local exposure, with larger declines for a county's first hyperscale opening and smaller declines in established clusters. Third, capital-

---

[1] Variables include disaster exposure and resilience, zoning and land-use regulation, water availability and rights, storm risk and damages, environmental regulation, power-generation and grid infrastructure, renewable energy capacity, broadband and fiber connectivity, utility reliability, electricity prices, climate conditions, county- and state-level subsidy policies, and tract-level demographics and employment characteristics.

[2] This aligns closely with industry accounts that large data centers increasingly seek sites with abundant land, reliable power, and fiber connectivity, often outside urban cores in suburban or lower-density areas (Cushman & Wakefield, 2025).

ization effects are concentrated within the same zoning jurisdiction as the facility and attenuate across jurisdictional boundaries, consistent with the regulatory context shaping perceived exposure. Taken together, these results indicate that local capitalization depends not only on burden intensity, but also on whether those burdens are introduced in settings where they are more exposed, and therefore more salient, to nearby homeowners.

We also examine additional incidence channels. We find little evidence of significant or sustained employment responses across multiple datasets and spatial scales, suggesting that offsetting local labor-market gains are unlikely to account for the observed price declines. By contrast, residential electricity costs rise modestly after hyperscale openings, consistent with higher household utility burdens. Evidence based on water-service-area controls further suggests that an important component of capitalization operates through highly localized utility-service conditions, although these tests are not separately definitive. We also find suggestive evidence that environmental household burdens vary with wind exposure, with larger price declines for parcels more frequently downwind of hyperscale sites.

This study contributes to several strands of literature. First, it contributes to the literature on place-based development and the local effects of large capital investments. Existing work on major plant openings and other large projects often emphasizes local gains through labor demand, productivity, and housing demand (Ahlfeldt et al., 2015; Allcott and Keniston, 2018; Busso et al., 2013; Giroud et al., 2024; Greenstone et al., 2010; Kline and Moretti, 2014b,a; Zheng et al., 2017). Our results highlight an important contrast: hyperscale data centers represent a different class of investment. They are highly capital-intensive, labor-light, and infrastructure-intensive, to which we find little evidence of offsetting local employment gains. In this setting, nearby housing markets capitalize primarily localized costs rather than broad local growth. More broadly, the paper shows that the local incidence of place-based investment depends not only on the project scale, but also on the production technology and operational footprint of the investment itself.

Second, the paper contributes to the emerging literature on digital infrastructure and its local economic effects (Benetton et al., 2023; Feher et al., 2025; Greenstein and Fang, 2022; Halaburda and Yermack, 2023; Knittel et al., 2025). Although data centers are now a foundational input into cloud computing and artificial intelligence, evidence on their local welfare consequences remains limited. We provide the first national-level causal evidence on how data center openings affect nearby residential property values, and show that the negative capitalization effect is concentrated among hyperscale facilities rather than data centers more generally. This distinction is central: it implies that the local effects of digital infrastructure depend critically on scale, resource intensity, and siting context. More broadly, the paper also speaks to the growing literature on the economic consequences of artificial intelligence by studying one of its core physical inputs. Whereas existing work focuses primarily on AI adoption at the firm, worker, or financial-decision level, our results show that the infrastructure supporting AI can generate localized costs in surrounding communities.[3]

Third, the paper relates to the literature on housing-market capitalization and the broader study of economic incidence (e.g., Banzhaf and Walsh, 2008; Black, 1999; Chay and Greenstone, 2005; Cellini et al., 2010; Davis, 2004; Linden and Rockoff, 2008; Roback, 1982; Rosen, 1974). We use housing prices to measure how the net local effects of hyperscale openings are distributed across space, and the estimates reveal a sharp spatial gradient: costs are concentrated near the facility, while any fiscal or broader economic gains are likely to accrue at wider geographic scales. More specifically, the paper extends the hedonic valuation literature on localized industrial and infrastructure disamenities by introducing hyperscale data centers as a distinct form of land use, one whose externalities arise not from traditional manufacturing activity but from resource-intensive digital operations and supporting infrastructure.[4] The paper thus identifies a form of spatial decoupling in which the benefits and burdens of investment need not be geographically aligned. This broader incidence insight has implications beyond data centers, including for the evaluation of local subsidies, infrastructure policy, and other capital-intensive projects whose benefits are diffuse but whose costs are borne by nearby residents.

The remainder of the paper proceeds as follows. Section 2 presents the conceptual framework. Sections 3 and 4 describe the data, sample construction, and empirical strategy. Section 5 presents the baseline results, while Section 6 examines mechanism evidence and incidence channels. Section 7

---

[3] Prior work finds 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 labor-market disruption (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), while AI adoption can reduce behavioral biases in trading decisions (e.g., D'Acunto et al., 2019) and improve loan underwriting performance, but may also increase disparities between minority and non-minority borrowers (e.g., Jansen et al., 2024; Fuster et al., 2022).

[4] Existing work has documented negative housing-market capitalization effects for a wide range 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).

provides a back-of-the-envelope calculation, and Section 8 concludes.

2 Conceptual Framework

We model a data center opening as a local investment shock whose benefits and costs may accrue at different geographic scales. Let utility at location l in jurisdiction c, for a facility of type s, be given by

$$U_{i,l,c,s} = W_{i,l,c,s} + G_{c,s} + A_{l,c,s} - K_{i,l,c,s} - R_{l,c,s}, \tag{1}$$

where $W_{i,l,c,s}$ denotes labor-market benefits, $G_{c,s}$ denotes jurisdiction-wide fiscal benefits, $A_{l,c,s}$ denotes location-specific residential amenities, $K_{i,l,c,s}$ denotes household costs such as utility expenses, $R_{l,c,s}$ denotes housing costs. In spatial equilibrium, changes in local utility are capitalized into housing prices. We therefore use housing prices to measure households' valuation of the net local effects of a nearby data center opening.

A data center may raise $G_{c,s}$ by expanding local public revenues, and it may raise $W_{i,l,c,s}$ through local labor demand, although the latter channel may be limited when operations are highly capital-intensive and require little ongoing labor. At the same time, it may reduce $A_{l,c,s}$ if operations generate localized disamenities and may raise $K_{i,l,c,s}$ if concentrated electricity and water demand increases residential utility costs through pass-through or infrastructure upgrades.[5] Because housing prices capitalize perceived local utility, the extent to which operational intensity translates into residential disamenity can depend on local exposure and salience. It is useful to write

$$A_{l,c,s} = \bar{A}_{l,c} - \sigma_{l,c,s} \cdot D_{l,c,s},$$

where $D_{l,c,s}$ denotes the underlying intensity of localized burdens generated by the facility's operations and supporting infrastructure, and $\sigma_{l,c,s}$ captures how strongly these burdens translate into perceived residential disamenity and thus into capitalization. These effects need not be uniform across facility types and locations. In our setting, facilities associated with large cloud and AI operators differ systematically from other data centers in both technical intensity and location

---

[5] Potential localized disamenities include persistent noise (e.g.,Cary, 2023; Carlson and Teplitzky, 1974; Ellis, 1971; Zhouo et al., 2020), heat (e.g., Chen et al., 2026), and other emissions (e.g., Han et al., 2025).

context: they are more resource-intensive, more often located in lower-density, residential-oriented environments, and more likely to be developed as campus-style sites, where an initial facility may be followed by later expansion. These distinctions are documented in the data and revisited in subsequent sections. These features can increase both $D_{l,c,s}$ and $\sigma_{l,c,s}$, amplifying local capitalization. By contrast, other facilities are smaller and more often embedded in existing commercial or institutional areas with greater nearby activity, where incremental burdens may be less salient to nearby households. As a result, localized disamenities and utility burdens may be more likely to dominate near hyperscale sites, even if jurisdiction-wide fiscal gains are positive.

Our empirical objective is to identify the localized incidence of data center openings on nearby households, holding broader jurisdiction-level effects fixed. We do so by comparing homes near a facility with homes in the same county that are slightly farther away. Under the assumption that broader labor-market and fiscal effects are shared within the relevant local area, these comparisons difference out much of the common variations in $W_{i,l,c,s}$ and $G_{c,s}$. The resulting estimates therefore capture the differential local effects operating through amenities and household costs—reflected in changes in $A_{l,c,s}$ and $K_{i,l,c,s}$—rather than the aggregate welfare effect of the investment for the broader jurisdiction.

3 Data

3.1 Data Sources and Sample Construction

Data Center Openings Database. We obtain facility-level information on U.S. data centers from the S&P Global 451 Research Knowledge Base. The database provides geographic coordinates, construction and operational years, facility type and size, technical capacity, and ownership information. Our final sample includes all data centers with non-missing geocoordinates and either a construction year or an operational year. When the construction year is missing, we use the operational year as the closest proxy.

Housing Transaction. We obtain housing transaction data from CoreLogic, a widely used commercial database of U.S. property sales compiled from public deed and assessor records. The data contain transaction prices and dates, together with a rich set of property and transaction characteristics. We restrict the sample to residential properties, including single-family homes, condominiums, duplexes, and apartments. We retain arm's-length transactions and exclude intra-family transfers. The analysis covers 2000–2025, and all sale prices are converted to real 2010 dollars.

Census and Auxiliary Data We combine the housing and facility data with several auxiliary datasets used for heterogeneity and incidence-channel analysis. To measure local demographic and socioeconomic conditions, we use the 2000 and 2010 Decennial Censuses together with the 2006–2019 American Community Survey (ACS) 5-year estimates. These data provide tract-level measures of population, income, racial composition, housing tenure, and utility expenditures.

To study local labor-market effects, we draw on two employment datasets. First, the Longitudinal Employer–Household Dynamics (LEHD) Origin–Destination Employment Statistics (LODES) provide annual block-level employment counts by workplace location and industry. These data allow us to measure both total employment and employment in sectors plausibly related to data center activity. Second, as an alternative source, we use county-level data from the Bureau of Labor Statistics Quarterly Census of Employment and Wages (QCEW), from which we construct total covered employment and average weekly wages by county-year and sector.

We also use additional data to examine specific channels. To characterize land use around data centers, we use panel land-use data at five-year intervals from the Historical Settlement Data Compilation for the U.S. (HISDAC-US; 1995–2020; Ahn et al., 2024). For each facility, we compute the pre-treatment share of each HISDAC land-use class within 5 km, where pre-treatment is defined as the most recent interval preceding the facility's year built.[6] To study water-system exposure, we use SimpleLab's TEMM (Tiered Explicit, Match, and Model) dataset, which provides nationwide community water-system service areas. We assign each geocoded housing transaction to a water-system boundary by spatial join. To study directional exposure to potential environmental costs, we use near-surface wind vectors from the North American Regional Reanalysis (NARR), obtained from the National Oceanic and Atmospheric Administration. For each facility site, we combine the site-level wind direction with the bearing from the facility to each parcel to construct monthly indicators of whether a property lies downwind or upwind of the site.

---

[6] For example, a facility built in 2014 is matched to the 2010 HISDAC-US layer.

Sample Construction. We spatially link data centers to housing transactions using the geocoordinates of both facilities and properties. Using these same coordinates, we assign each transaction to its corresponding census tracts and blocks and merge in additional spatial and auxiliary data via geographic overlays or shared census identifiers.

For each data center, we construct an event-level sample consisting of all housing transactions within a 50 km radius of the site. In the baseline sample, properties within 5 km form the near ring, while properties 20–50 km away form the outer ring; the intermediate 5–20 km band is excluded to limit potential spillovers into the comparison group. We then stack these facility-level subsamples into a single estimation dataset. If a property is exposed to multiple data centers over time, the exposure is assigned to the first facility. We also exclude from the control group any transaction within 20 km of any data center.

3.2 Summary Statistics

Figure 1 summarizes the growth of data center construction in the United States over time. More than 500 facilities were built before 2010, and construction increases steadily over the sample period. Hyperscale facilities—defined here as data centers built, owned, and operated by major cloud and technology firms (Amazon, Apple, Google, Meta, Microsoft)—account for a relatively small share of activity before 2010, but their construction accelerates sharply after 2020. By 2025, the number of newly built hyperscale facilities approaches that of all other data center types combined.

Table 1 compares facility characteristics and local conditions around hyperscale and non-hyperscale sites in the estimation sample. Hyperscale facilities account for roughly 15% of the sample, while the non-hyperscale sample is dominated by retail and wholesale facilities. Relative to non-hyperscale sites, hyperscale facilities are substantially larger and more power-intensive, with greater floor area, more racks, and higher power per rack.

The surrounding areas differ systematically along demographic, land-use, and workplace dimensions. Hyperscale sites are located in lower-density environments: tract-level data show lower population density (0.291 versus 1.111 persons per km^2) and housing-unit density (0.110 versus 0.546), but higher owner-occupancy (0.661 versus 0.474). At the same time, pre-treatment land use surrounding hyperscale facilities is substantially more residential and less oriented toward commercial, industrial, and government uses. The mean residential land share is 0.65 around hyperscale sites, compared with 0.48 around non-hyperscale sites, while the corresponding commercial/industrial/government share is 0.29 versus 0.45. These differences indicate that hyperscale facilities are not simply larger data centers; they are also sited in environments that are systematically more residentially exposed. By contrast, non-hyperscale sites are located in more employment-intensive areas, with substantially higher nearby workplace jobs overall—and especially in business and consumer services—consistent with their being more embedded in existing economic activity centers.[7]

Table 2 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. For the full sample of data centers, transactions near sites have higher average sale prices, but the underlying properties are smaller, older, and located in neighborhoods with a higher Black share, lower median income, and higher unemployment. By contrast, for hyperscale facilities, observable differences across distance bands are much smaller. Mean prices are slightly lower in the near ring, while property and neighborhood characteristics are broadly similar between the near and outer rings.

4 Empirical Strategy

The main empirical concern is that data center openings are not randomly assigned. Developers choose locations based on economic, infrastructural, regulatory, and political factors, many of which may also correlate with underlying housing-market trends. A simple comparison of prices near and far from data centers would therefore confound the effect of opening with non-random siting. Our objective is to identify the *localized* incidence of a facility opening on nearby households while holding broader local shocks as fixed as possible.

Our baseline specification is a stacked difference-in-differences design that compares price changes in the near ring (≤ 5 km) with those in the outer ring (20–50 km) around the same facility. The specification includes a rich set of property- and transaction-level controls, data-center-by-distance

---

[7] Supplementary evidence from the FEMA–ORNL–USGS USA Structures database points in the same direction: non-hyperscale facilities are located near substantially more commercial and institutional structures (Appendix Table A-1). Because the FEMA data are cross-sectional, we use them as a complement to the panel land-use measures from Ahn et al., 2024, which are available at five-year intervals from 1980 and allow us to characterize the pre-treatment land-use context.

fixed effects, tract-by-cohort fixed effects, and county-by-year-month-by-cohort fixed effects.[8] This comparison differences out shocks common to the surrounding local housing market, including county-level changes in housing demand and supply, macroeconomic conditions, and other time-varying factors shared across nearby locations. To the extent that broader fiscal or labor-market effects are within the local area, it also nets out part of the jurisdiction-wide component of the opening.

A remaining concern is that, even within the same county, locations near eventual data center sites may differ systematically from locations farther away. To strengthen comparability, we therefore replace the outer-ring comparison group with the near ring (≤ 5 km) around facilities that open later. Because both groups consist of locations similarly close to sites eventually selected for treatment, this design differences out more of the underlying siting selection process and identifies the effect from differences in treatment timing across eventually treated locations. The identifying assumption is that, absent the opening, prices near currently treated sites would have evolved similarly to prices near later-treated sites, conditional on controls and fixed effects.

To further strengthen comparability, we estimate siting propensity scores using XGBoost, an efficient implementation of gradient boosted decision trees (GBDT), with a broad set of pre-treatment predictors at both the county and tract levels.[9], [10] County covariates capture infrastructure, resource constraints, and policy environments, while tract covariates capture local demographics and economic conditions.[11] We then match each treated near ring to later-treated locations with similar predicted siting probabilities and use these matched later-treated near rings as the comparison group. This design restricts attention to locations that were observably similar ex ante and

---

[8] The intermediate 5–20 km band is excluded in the baseline specification to reduce the scope for spillovers into the comparison group. We recognize that the choice of distance bands is somewhat arbitrary. Accordingly, later sections examine distance-gradient effects using a series of finer, non-overlapping distance bins (e.g., 0–1 km, 1–2 km, . . . , up to 19–20 km), and report robustness checks that vary the distance cutoffs and reintroduce the intermediate band.

[9] For applications of GBDT in economics, see, for example, Kleinberg et al., 2018 and Gu et al., 2020; for a recent application of XGBoost, see Ash et al., 2025.

[10] GBDT combines many shallow trees to capture nonlinearities and interactions in county and tract characteristics. To impose an economically sensible structure, we constrain a small subset of predictors to enter monotonically in theoretically motivated directions—for example, higher industrial electricity prices cannot increase predicted siting probability; other constrained predictors include disaster exposure, National Ambient Air Quality Standards (NAAQS) nonattainment status, substation capacity, high-voltage transmission access, and fiber infrastructure. The resulting constrained model performs well out of sample, achieving an ROC AUC of 0.731 on the held-out test set.

[11] County-level predictors include 2006–2010 averages of measures of disaster risk, water conditions, environmental regulation, energy and transmission infrastructure, renewable capacity, fiber connectivity, climate, and local subsidy policy. Tract-level demographic and housing predictors are drawn from the 2006–2010 ACS 5-year estimates, while tract-level employment measures are constructed from 2006–2010 averages of LODES data.

also at risk of treatment, so identification comes from differences in opening timing among more comparable future-treated sites.

Our specification is as follows:

$$y_{i,t,j,c,d} = \beta_1 \cdot \mathbf{1}\{\text{Dist}_{id} \leq 5km\} \times \mathbf{1}\{\text{Post}_{t,d} = 1\} + X'_{i,t}\Gamma + \alpha_{j,d} + \theta_{c,t,d} + \epsilon_{i,t,j,c,d}, \tag{2}$$

where $y_{i,t,j,c,d}$ is the log transaction price of property i, located in census tract j and county c, transacted in year-month t, in data center cohort d. $\text{Dist}_{id} \leq 5km$ indicates that the property lies in the treatment area for opening d, and Post_{td} indicates periods after that opening. In the baseline specification, the exposed area is the near ring around the facility and the comparison group is the outer ring around the same site. In the later-treated and matched later-treated designs, the exposed area remains the near ring around the current opening, but the comparison group is instead the near ring around sites that open later, with the latter restricted to matched sites in the most demanding specification. X_{it} includes a rich set of property- and transaction-level controls, α_{jd} are tract-by-cohort fixed effects, and θ_{ctd} are county-by-year-month-by-cohort fixed effects. β therefore measures the differential change in prices in the exposed area after opening, relative to the relevant comparison group.

5 Main Results

5.1 Impact on Housing Prices

Table 3 reports baseline stacked difference-in-differences estimates of housing-price responses to data center openings. We report specifications with progressively richer property- and transaction-level controls and increasingly granular time fixed effects, with our preferred specification in Column 6 including county-by-year-by-month fixed effects that absorb county-level shocks common to both distance rings around the same data center.[12]

Panel A, which pools all data center openings, yields economically small and imprecise estimates.

---

[12] Property- and transaction-level controls include housing characteristics (e.g., property type, size, age, bedrooms, bathrooms, structural attributes such as type of floor and roof), and indicators for cash purchases, resales, distressed sales (short sale and foreclosure/REO), and other transaction circumstances, including buyer and seller characteristics, where available. We also include flexible distance-to-facility controls (1-km distance-bin fixed effects) and absorb local and time-varying shocks using tract fixed effects and increasingly granular time fixed effects.

This pooled null is consistent with the heterogeneity documented in Section 3.2: the full sample is dominated by smaller retail and wholesale facilities whose operational footprint is unlikely to generate detectable localized costs at a 5 km radius. Before turning to hyperscale results, it is useful to clarify what distinguishes these facilities from the broader data center stock. As documented in Table 1, hyperscale facilities—those built, owned, and operated by major cloud and AI firms, including Amazon, Google, Microsoft, Meta, and Apple—are substantially larger and have higher power capacity than other data centers, with nearly five times as many racks and much higher power per rack. Critically, they are more often sited in lower-density, residential-oriented environments where operational burdens are more salient to nearby homeowners, and they are more likely to develop as campus-style projects that generate sustained rather than transient disruption. In the conceptual framework, these features imply both higher burden intensity $D_{l,c,s}$ and greater salience $\sigma_{l,c,s}$ —precisely the conditions under which localized disamenities are most likely to outweigh any jurisdiction-wide fiscal gains.

Panel B, which focuses on hyperscale facilities, shows a clear negative capitalization effect. Transaction prices within 5 km fall by 6.8% relative to the 20–50 km ring after a hyperscale opening (Column 6).This magnitude is economically meaningful and comparable to capitalization effects documented for other prominent locally undesirable facilities: Davis (2011) finds a 3%-7% reduction in housing values near newly opened fossil-fuel power plants, and Muehlenbachs et al. (2015) document declines of similar magnitude near shale gas development. The estimated price effects are stable across specifications, suggesting that these estimates are unlikely to reflect state- or county-level housing cycles or compositional shifts in transacting properties. Appendix Table A-2 further shows that the estimate is robust to alternative comparison rings: using a closer control ring (5–50 km) attenuates the estimate to -3.4%, while restricting controls to more distant rings yields effects similar to baseline, ranging from -5.6% to -8.9%. Taken together, the contrast between Panels A and B indicates that the negative capitalization effect is concentrated among hyperscale facilities rather than data centers more generally.[13]

Figure 2 plots event-study estimates for hyperscale openings, normalized to the year immediately

---

[13] One possible explanation for the muted full-sample estimates is that smaller facilities generate more localized spillovers, so a 5 km treatment radius may attenuate impacts. Appendix Table A-3 restricts the sample to non-hyperscale facilities and redefines treatment using narrower distance bands. The estimates remain economically small and statistically indistinguishable from zero.

before opening. The pre-opening coefficients are small and statistically indistinguishable from zero, providing no evidence of differential pre-trends between the near and outer rings. After opening, the estimates become negative by about year 2 and remain below zero thereafter, settling at a persistent decline of approximately -5%.[14] Appendix Figure A-2 shows that a consistent dynamic pattern obtains under alternative staggered-adoption estimators designed to address concerns with two-way fixed-effects specifications in settings with staggered timing (Borusyak et al., 2024; Callaway and Sant'Anna, 2021; De Chaisemartin and d'Haultfoeuille, 2024; Sun and Abraham, 2021). In addition, we re-index treatment time to the earliest documented project milestone; as shown in Appendix Figure A-3, there is little evidence of an immediate response at early stages of development.[15] Taken together, these timing patterns are less consistent with a purely contemporaneous visual disamenity and more consistent with burdens that emerge during operations or with gradual learning about local externalities.

To directly address the concern that the near ring may differ systematically from the outer ring, we replace the outer-ring comparison group with a "later-treated" near ring—transactions within 5 km of facilities that open later—so that treated and control groups are drawn from neighborhoods close to sites ultimately selected for hyperscale facilities. The upper panel in Figure 3 shows that the post-opening coefficients are consistently negative and remain below zero in subsequent years, consistent with the baseline pattern. We then implement a matched later-treated design that pairs each current opening with five later-treated sites that have the closest pre-treatment siting probability, where the predicted probability is estimated using an XGBoost classifier, a gradient-boosted tree model, over a rich set of county- and tract-level predictors. As shown in the lower panel, the resulting estimate remains negative and similar in magnitude, suggesting that the baseline finding is not driven by the particular choice of the outer-ring comparison group.[16]

---

[14] A delayed response could reflect subsequent nearby data center entries. Appendix Figure A-1 re-estimates the model after excluding transactions that occur in the year of, or after, any subsequent opening; the decline remains and is slightly larger.

[15] The alternative timing variable is defined as the earliest of three project milestones identified from publicly available sources: the first public announcement year, the initial government-engagement year, and the construction-start year. We collected these dates using project names, company names, and site addresses in ChatGPT-assisted searches of news coverage, planning documents, government records, and company materials. Because public documentation is incomplete and project names are not always unique, the resulting milestone years are measured with error and should be interpreted as approximate timing proxies.

[16] We use five nearest-neighbor matches for each treated tract based on XGBoost-predicted siting probabilities. Appendix Figure A-4 shows a consistent dynamic pattern when using either 10 or 3 nearest-neighbor matches.

5.2 Distance Gradient in Price Capitalization

We next examine how capitalization varies with proximity by allowing treatment effects to differ flexibly across distance bins. Specifically, we re-estimate the stacked difference-in-differences specification interacting the post-opening indicator with non-overlapping distance rings (0–1 km, 1–2 km, . . . , 19–20 km), using transactions in the 20–50 km ring as the reference group. This approach provides a transparent characterization of spatial incidence and helps delineate the geographic scope over which nearby housing markets respond.

Figure 4 shows a steep distance gradient in housing-price responses. Effects are largest closest to the facility—about 15% within 1 km—and attenuate with distance, falling to roughly 9% within 1–4 km and remaining economically meaningful through intermediate distances before becoming small and statistically indistinguishable from zero beyond about 14 km.[17] This pattern indicates that the average 0–5 km estimate masks substantial heterogeneity, with the negative capitalization concentrated among households living very close to the site. At the same time, the persistence of effects well beyond the immediate vicinity is suggestive of a mechanism broader than a purely visual disamenity operating only through line-of-sight exposure. Relative to the housing-capitalization evidence for other industrial disamenities, the spatial reach here is unusually broad: Davis, 2011 finds 3–7% declines within roughly 2 miles of power plants; Currie et al., 2015 find an 11% decline within 0.5 mile of toxic plant openings, with air-quality effects dissipating within 1 mile; Muehlenbachs et al., 2015 find the strongest negative effects from shale-gas development within 1–2 km. This implies that the local incidence of hyperscale development is not confined to immediate adjacency, which is directly relevant for siting decisions, buffer-zone design, and the evaluation of place-based incentives that may generate diffuse benefits but localized housing-market costs (e.g., Hamilton and Schwann, 1995; Zou, 2020).

6 Incidence Channels

Guided by the conceptual framework (Equation 1), this section examines the mechanisms underlying the localized housing-market effects documented above. We organize the evidence around four

---

[17] Precision is lower in the innermost bin because relatively few transactions (slightly above 2,000) occur within 1 km of hyperscale facilities.

margins: variation in the intensity of localized burdens generated by hyperscale facilities, differences in how those burdens are capitalized into housing prices, direct household cost channels, and the extent of offsetting local labor-market benefits.

6.1 Operational Capacity

We begin by examining whether facilities with higher power capacity generate larger localized housing-market effects. In the conceptual framework, facilities with greater power capacity are likely to impose larger localized burdens, implying larger capitalization effects where power capacity is higher.

We first examine heterogeneity by power capacity within hyperscale facilities.[18] Table 4 shows a clear dose-response pattern: restricting the sample to facilities with at least 100 MW of UPS power yields a large and precisely estimated post-opening decline of −24%, while lowering the threshold to 80 MW, 60 MW, and 40 MW yields smaller declines of −8.8%, −9.2%, and −6.4%, respectively. This gradient indicates that the local housing-market response is increasing in power capacity within the hyperscale segment.

An important question is whether scale alone is sufficient to generate negative capitalization outside the hyperscale segment. Appendix Table A-4 reports analogous estimates for large non-hyperscale wholesale facilities. Even at high capacity thresholds, estimated price effects remain small and statistically indistinguishable from zero. This contrast suggests that the incidence of large facilities depends on more than power capacity alone.

6.2 Salience and Local Exposure

We next examine how local context shapes the salience with which operational burdens are capitalized into housing prices. To understand why large wholesale facilities exhibit little capitalization on average, we begin with differences in surrounding land-use context. Figure 6 revisits the land-use contrasts documented in the summary statistics, restricting attention here to large facilities only. Hyperscale sites are substantially more residentially exposed: the median residential share around

---

[18] We proxy operational intensity using facility-level Total Uninterruptible Power Supply (UPS). 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 the IT load supported by the facility.

hyperscale sites is 0.75, compared with 0.45 around large wholesale sites. The reverse pattern holds for commercial, industrial, and government land, whose median share is 0.18 around hyperscale sites versus 0.44 around large wholesale sites. These figures indicate that the two types of large facilities are embedded in systematically different local environments, implying that similar operational burdens are likely to be more salient to nearby homeowners around hyperscale facilities.

We test this implication directly within the sample of large non-hyperscale facilities. Appendix Table A-5 shows that the average null effect for large sites masks substantial heterogeneity by surrounding land use. When large non-hyperscale facilities are located in the lowest quartile of nearby commercial, industrial, and government land share, nearby house prices decline by 8.9 percent after opening. This magnitude is economically meaningful and comparable to the baseline hyperscale estimate of 6.8 percent. By contrast, effects are statistically insignificant in more commercially embedded settings. Taken together, these results indicate that operational capacity alone is not sufficient to generate negative capitalization; the same large facility can have very different local incidence depending on whether it is sited in a residentially exposed or commercially embedded environment.

To further assess whether the results are consistent with a role for salience, we examine two complementary dimensions. First, we use Google Search intensity as a proxy for local public attention around facility openings.[19] Figure **??** shows little evidence of responses in search activity prior to opening, but a pronounced post-opening increase thereafter. The response is largest for hyperscale facilities, rises more moderately for wholesale facilities, and remains close to zero for retail sites. This descriptive pattern suggests that hyperscale openings are substantially more salient in local information environments, consistent with their larger scale and greater visibility to nearby communities.

We then turn to a second measure of salience: prior local exposure to hyperscale development. Appendix Table A-6 shows that the estimated decline is largest for a county's first hyperscale opening and attenuates for subsequent openings; effects are also substantially larger in counties with low overall data-center presence than in established clusters. This pattern is consistent with the

---

[19] Google Search Index is sourced from Google Trends, which normalizes search intensity for a given term to a scale of 0 to 100 relative to peak search volume within a region and time period. We download monthly search intensity for the term "data center" for all 50 U.S. states over 2004–2026, and align each of the facilities in our sample to its respective state's index series using the facility opening year as T=0, constructing an event window from T-4 to T+4. The index captures within-state variation in public attention over time.

idea that the same operational burdens are capitalized more strongly when hyperscale development is novel to a local area and therefore represents a larger perceived change in local conditions. While these results are suggestive rather than definitive, they indicate that local capitalization depends not only on burden intensity, but also on the spatial and temporal context in which those burdens are introduced.

6.3 Jurisdiction Boundaries and Regulatory Exposure

We further investigate whether localized capitalization depends only on physical proximity or also on shared regulatory jurisdiction. If housing-price effects are driven solely by physical disamenities such as noise, heat, or airborne emissions, they should vary smoothly with distance from the facility, rather than discontinuously at administrative borders. By contrast, if households also capitalize expectations about future land-use exposure, permitting decisions, or expansion risk, similarly proximate homes may be affected differently depending on whether they fall within the same jurisdiction as the data center.

To test this, we distinguish nearby homes within 5 km of a hyperscale facility according to whether they lie inside the same zoning jurisdiction as the data center site or across a zoning boundary in an adjacent jurisdiction.[20] Table 5 shows a clear asymmetry: the negative capitalization effect is concentrated among homes located within the same zoning jurisdiction as the facility, while nearby homes across jurisdiction boundaries exhibit little or no statistically detectable decline.

This pattern is difficult to reconcile with a purely distance-based physical disamenity story, under which similarly proximate homes should experience comparable effects regardless of administrative boundaries. Instead, it suggests that shared regulatory jurisdiction shapes how nearby households assess the local consequences of hyperscale development. Homes within the same jurisdiction may face greater expected exposure to future expansion, land-use changes, or permitting spillovers, leading them to capitalize hyperscale openings more strongly into housing prices.

---

[20] We obtain zoning-jurisdiction boundaries from the NZA (Bronin and Derickson., 2026); where this information is missing, we use county boundaries instead in later regressions as a robustness check.

6.4 Utility Costs

In this section, we examine direct household cost channels through which hyperscale operations may affect nearby residents. Hyperscale facilities add large, geographically concentrated electricity loads that can require grid investments and raise procurement costs; to the extent that these costs are recovered through retail rates, households may face higher utility bills.

We study utility costs using census-tract measures of average annual household electricity and water costs from the ACS and an analogous tract-level event-study design.[21] Figure 8 shows a modest but persistent increase in electricity costs following hyperscale commissioning: post-opening coefficients become positive and remain elevated at approximately 2%.[22] By contrast, water-cost responses in the ACS are economically small and statistically indistinguishable from zero, with no sustained post-opening increase. Because retail electricity rates and many grid costs are determined at broader service-area or jurisdictional levels, we interpret this pattern as evidence of higher household costs, though not as a complete explanation for the sharp within-county distance gradient in housing prices.

Given concerns that water-related burdens may operate at the level of the water utility rather than through tract-average billing measures, we next assess whether the housing-price estimates are sensitive to absorbing water-service-area-by-time variation. Using water-service boundaries from SimpleLab TEMM, Appendix Table A-7 augments the baseline housing-price specification with flexible water-system-by-time fixed effects, which absorb shocks common to households served by the same utility. Relative to the baseline estimate, the housing-price effect attenuates substantially and becomes statistically weak when these fixed effects are included. This attenuation persists when we retain unmatched transactions, using a missing-category indicator as the boundary identifier.

We interpret this sensitivity cautiously because water-system-by-time fixed effects absorb multiple highly localized channels simultaneously. On one hand, the attenuation is consistent with shocks operating at the utility-service level, including potential water stress, infrastructure upgrades, or

---

[21] Because the ACS measures electricity and water costs at the census-tract level, we define treatment and comparison groups at the tract level. Treated tracts are those whose centroids lie within 5 km of a hyperscale opening, and comparison tracts are those 20–50 km from the same facility. Controls 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.

[22] Results are 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).

reliability changes shared within a water-system boundary.[23] On the other hand, water-service boundaries may also absorb other place-specific factors correlated with siting—including localized amenities and disamenities and information or expectation shocks that are common within service areas. These specifications, therefore, do not isolate water stress per se; rather, they indicate that an important component of capitalization operates through highly localized factors that covary at the utility-service-area level.

6.5 Environmental Disamenity

Beyond direct monetary costs, hyperscale operations may also impose environmental burdens that vary across households depending on wind exposure. If hyperscale operations generate airborne emissions or wind-dependent propagation of mechanical noise, impacts should be larger for locations that are more frequently downwind of the facility (Anderson, 2020; Briggs, 1975; Deryugina et al., 2019; Embleton, 1996; International Organization for Standardization, 2024; Ruiz et al., 2016). We test this implication using a parcel-level downwind-propensity measure constructed from long-run prevailing winds and estimate a triple-difference specification that interacts the post-opening effect in the near ring with downwind propensity.[24] The estimated interaction is negative but imprecisely estimated. Extreme-subsample comparisons point in the expected direction—price declines are larger for always-downwind parcels (-12.5%) than for never-downwind parcels (-0.9%)—though these estimates are likewise noisy and potentially sensitive to compositional differences. Appendix Table A-8 reports the full results.

6.6 Limited Local Labor-Market Benefits

State and local governments commonly offer tax incentives to attract data center projects, citing large capital investment, expansion of the local tax base, and broader economic-development spillovers.[25] Hyperscale facilities, however, are highly automated and typically employ relatively

---

[23] If hyperscale operations tighten local water availability, require system upgrades, or affect reliability and service quality in ways shared within a water-system boundary, then absorbing water-system-specific time variation would remove precisely the component of capitalization operating through water-utility conditions.

[24] We construct downwind propensity as 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. "Always downwind" and "never downwind" refer to the two extremes of this measure.

[25] The state offered incentives tied to capital investment and job creation. See, Amazon to invest $11 billion in Indiana to build data centers.

few workers once operational; as such, local labor-market gains may be limited relative to the size of the investment.[26] Here, we examine whether hyperscale openings generate localized employment or earnings gains; because our housing-price estimates compare the near ring to a farther ring within the same county and month, the relevant question is whether labor-market outcomes change differentially in the immediate vicinity of an opening.

Figure 7 reports event-study estimates using census block-level LODES Workplace Area Characteristics (WAC) data, focusing on sectors most plausibly linked to hyperscale operations and related activity (Utilities, Information, and Professional, Scientific, and Technical Services). The estimates show little evidence of sustained post-opening employment gains in near-ring blocks relative to outer-ring blocks. At longer horizons, we observe at most a modest increase in the Professional, Scientific, and Technical Services sector, while estimates for Utilities and Information remain close to zero. Overall, the evidence suggests muted localized labor-demand effects.[27]

7 Back-of-Envelope Calculation

We interpret the estimated housing-price response as the present value of households' revealed willingness to pay to avoid the bundle of localized burdens associated with hyperscale operations. These burdens may include direct household costs, environmental disamenities, and changes in surrounding land-use conditions as perceived by nearby residents. Let H denote the aggregate capitalization effect implied by our preferred estimate. Using the pooled effect of 6.8%, an average sale price of \$350,000, and approximately 150,000 affected properties within 5 km of hyperscale facilities, we estimate $H \approx -\$3.6$ billion in our sample, or roughly $-\$23{,}800$ per affected property.

Under the standard hedonic interpretation (Bishop et al., 2020; Sheppard, 1999), this capitalization measure already reflects the net present value of the localized costs and benefits perceived by nearby households. This logic is important for interpreting our utility-cost results. Let E denote the present discounted increase in residential utility expenditures induced by hyperscale entry. Higher utility bills are, in principle, a negative welfare shock for local households, but they should not be added mechanically to H. To the extent that households anticipate these recurring costs and

---

[26] The AI boom has sharply raised the fiscal stakes of data-center subsidies. See, AI Boom Should Prompt States to Rein in Data Center Tax Losses.

[27] Appendix Figure A-5 provides complementary evidence using earnings bins, alternative spatial aggregations (county and commuting zone), and alternative data sources (QCEW); results are consistent.

incorporate them into housing demand, E is one component of the local amenity bundle already capitalized into house prices. Put differently, if $\kappa \in [0, 1]$ denotes the share of utility-cost increases already reflected in housing values, the benchmark hedonic case is $\kappa = 1$. We therefore interpret the utility estimates primarily as evidence on incidence channels that helps explain why $H < 0$, rather than as a separate term in our baseline incidence calculation.[28]

A related but conceptually distinct issue is fiscal incidence. Let S denote public subsidies committed to data center development. In our subsidy data, which span 32 counties containing more than 1,000 data centers, official commitments total at least \$16.1 billion (2024 USD). These subsidies are economically important, but they should not be interpreted as a direct social cost and added one-for-one to H. In general, subsidy payments are transfers from taxpayers to firms rather than mere capital losses. They are nevertheless relevant for local incidence. To the extent that a share $\phi \in [0, 1]$ of subsidy costs is borne by local residents through higher taxes, reduced public services, or foregone fiscal capacity, they represent an additional burden not necessarily captured by nearby housing capitalization. A broader accounting decomposition of localized welfare that fits this paper is therefore $W^{local} = H + \phi S + (1 - \kappa)E$. Our preferred revealed-preference estimate of localized residential incidence is thus $H \approx -\$3.6$ billion, while the broader expression clarifies why this figure should be interpreted as a lower bound once complementary fiscal- and cost-incidence channels are taken into account.

8 Conclusion

This paper provides the first national-level causal evidence on the local incidence of hyperscale data centers. Using stacked difference-in-differences designs applied to a national sample of data center openings, we find that proximity to hyperscale facilities is associated with economically meaningful declines in residential property values. In our preferred specification, housing prices within 5 km of a new hyperscale data center decline by approximately 6.8%, with effects that are steeply localized and become muted beyond roughly 14 km. This magnitude is comparable to capitalization effects documented for other prominent locally undesirable facilities. For instance,

---

[28] A separate add-on would be warranted only under incomplete capitalization—for example, if households do not fully observe or correctly forecast the utility-cost consequences of nearby hyperscale entry, so that $\kappa < 1$. We do not attempt to estimate that wedge here.

Davis (2011) documented a $3-7\%$ reduction in housing values within 2 miles of newly opened fossil-fuel power plants. Event-time estimates show no differential pre-trends and a post-opening decline that emerges with a lag, becoming negative about two years after completion and remaining below zero thereafter. The capitalization effect is stable across specifications and comparison groups, and it increases sharply with power capacity, with the largest impacts concentrated among the most power-intensive hyperscale sites.

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 localized employment or earnings in sectors most plausibly linked to data center operations. We also find that capitalization effects attenuate across zoning-jurisdiction boundaries, suggesting that shared regulatory context shapes how nearby households perceive and price hyperscale development. In contrast, we document a modest and persistent increase in residential electricity costs in host areas after commissioning, consistent with load-driven procurement costs and/or grid investments 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 substantially attenuates the estimated capitalization effect, which is consistent with channels operating at the utility-service level (including potential water-system conditions), although these specifications may also absorb other highly localized factors correlated with siting and neighborhood amenities. Finally, wind-direction tests yield negative but statistically imprecise heterogeneity by downwind propensity; while extreme-subsample comparisons point in the expected direction, these estimates remain noisy and suggestive rather than definitive.

A central implication of these results is that the local benefits and costs of hyperscale development are not spatially aligned. Because the estimated impacts decay sharply with distance, the localized burdens of hyperscale facilities, and the way those burdens are perceived within surrounding regulatory and land-use contexts, are capitalized into nearby housing markets, while fiscal benefits from development deals and tax-base expansion are likely to accrue over broader jurisdictions. 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 underscores the magnitude of these localized costs. Under the standard hedonic interpretation, the change in nearby house prices captures households' willingness to pay to avoid the bundle of localized burdens associated with hyperscale commissioning, net of any local benefits they value (Bishop et al., 2020; Sheppard, 1999). Applying our preferred estimate to the affected properties in our sample yields an aggregate capitalization loss of roughly $3.6 billion. As discussed in Section 7, this figure is best interpreted as a revealed-preference measure of localized residential incidence, with broader fiscal-incidence considerations potentially increasing the total burden borne by local residents. Consequently, local opposition to data center siting—often characterized as "NIMBYism"—may reflect the rational responses to localized burdens rather than generalized resistance to technological development.

Lastly, our analysis opens new avenues for future research. First, our donut-shaped design identifies localized effects net of broader county-level shocks shared by treated and comparison areas; it does not speak to general-equilibrium benefits of digital infrastructure that accrue outside the near ring. Second, future work could more directly measure the incidence of infrastructure upgrade costs by linking openings to interconnection agreements, utility tariffs, and grid investments. 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 calculations. 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 contributes by establishing a clear baseline: under current technological and policy conditions, hyperscale cloud infrastructure imposes measurable localized costs that are capitalized into nearby 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.

Ahlfeldt, Gabriel M, Stephen J Redding, Daniel M Sturm, and Nikolaus Wolf. 2015. "Hedonic analysis of housing markets." *Econometrica* 83 (6): 2127–2189.

Ahn, Yoonjung, Stefan Leyk, and Caitlin M. Mc Shane. 2024. "Historical Land Use for the U.S. 1940-2020: Major Class Version II." 10.7910/DVN/CHLNQG.

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.

Ash, Elliott, Sergio Galletta, and Tommaso Giommoni. 2025. "A Machine Learning Approach to Analyze and Support Anticorruption Policy." *American Economic Journal: Economic Policy* 17 (2): 162—93.

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.

Banzhaf, Spencer H, and Randall P Walsh. 2008. "Do people vote with their feet? an empirical test of tiebout's mechanism." *The American Economic Review* 98 (3): 843–863.

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.

Benetton, Matteo, Giovanni Compiani, and Adair Morse. 2023. "When cryptomining comes to town: High electricity-use spillovers to the local economy."Technical report, National Bureau of Economic Research.

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.*

Black, Sandra E. 1999. "Do better schools matter? parental valuation of elementary education." *The Quarterly Journal of Economics* 144 (2): 577–599.

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.

Bronin, Scott Markley Aline Fader, Sara C., and Evan Derickson. 2026. "National Zoning Atlas." `https://doi.org/10.4079/zoning-atlas`.

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.

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

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.

Cellini, Stephanie Riegg, Fernando Ferreira, and Jesse Rothstein. 2010. "The value of school facility investments: evidence from a dynamic regression discontinuity design." *The Quarterly Journal of Economics* 125 (1): 215—261.

Chay, Kenneth Y, and Michael Greenstone. 2005. "Does air quality matter? evidence from the housing market." *Journal of Political Economy* 113 (2): 376–424.

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

Chen, Yifan, Liu Ee Chia, Mingxuan Fang, and Qiang Wang. 2026. "When AI Gets Hot: The Impact of Data Centers on Microclimate." *Working Paper*.

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. 2004. "The effect of health risk on housing values: evidence from a cancer cluster." *American Economic Review* 94 (5): 1693—-1704.

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.

Giroud, Xavier, Simone Lenzu, Quinn Maingi, and Holger Mueller. 2024. "Propagation and amplification of local productivity spillovers." *Econometrica* 92 (5): 1589–1619.

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.

Greenstein, Shane, and Tommy Pan Fang. 2022. "Where the cloud rests: The location strategies of data centers." *Harvard Business School Working Paper 21-042.*

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.

Greenstone, Michael, Richard Hornbeck, and Enrico Moretti. 2010. “Identifying agglomeration spillovers: Evidence from winners and losers of large plant openings.” *Journal of Political Economy* 118 (3): 536–598.

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

Gu, Shihao, Bryan Kelly, and Dacheng Xiu. 2020. “Empirical Asset Pricing via Machine Learning.” *The Review of Financial Studies* 33 (5): 2223—2273.

Halaburda, Hanna, and David Yermack. 2023. “Bitcoin mining meets Wall Street: A study of publicly traded crypto mining companies.” *Working Paper, National Bureau of Economic Research.*

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

Han, Yuelin, Zhifeng Wu, Pengfei Li, Adam Wierman, and Shaolei Ren. 2025. “The Unpaid Toll: Quantifying and Addressing the Public Health Impact of Data Centers.” https://arxiv.org/abs/2412.06288.

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.

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 Quang Nguyen, and Amin Shams. 2024. “Rise of the Machines: The Impact of Automated Underwriting.” *Management Science* 71 (2): 955–975.

Kleinberg, Jon, Himabindu Lakkaraju, Jure Leskovec, Jens Ludwig, and Sendhil Mullainathan. 2018. “Human Decisions and Machine Predictions.” *The Quarterly Journal of Economics* 133 (1): 237—293.

Kline, Patrick, and Enrico Moretti. 2014a. “Local economic development, agglomeration economies, and the big push: 100 years of evidence from the Tennessee Valley Authority.” *The Quarterly Journal of Economics* 129 (1): 275–331.

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

Knittel, Christopher R, Juan Ramon L Senga, and Shen Wang. 2025. “Flexible data centers and the grid: Lower costs, higher emissins?” *Working Paper, National Bureau of Economic Research.*

Linden, Leigh, and Jonah E Rockoff. 2008. "Estimates of the impact of crime risk on property values from megan's laws." *American Economic Review* 98 (3): 1103—1127.

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

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

Roback, Jennifer. 1982. "Wages, rents, and the quality of life." *Journal of Political Economy* 90 (6): 1257–1278.

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

Rosen, Sherwin. 1974. "Hedonic prices and implicit markets: Product differentiation in pure competition." *Journal of Political Economy* 82 (1): 34–55.

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."

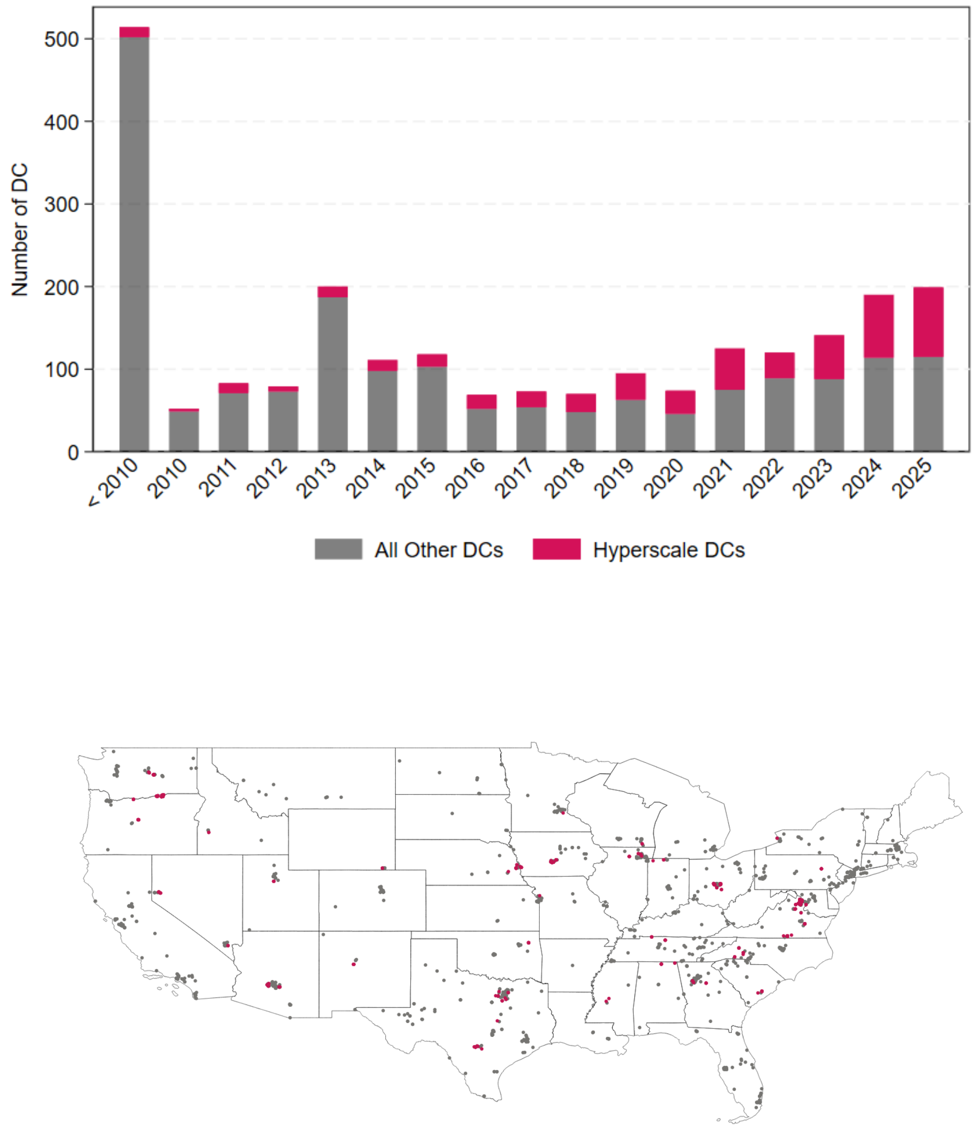
Zheng, Siqi, Weizeng Sun, Jianfeng Wu, and Matthew E Kahn. 2017. "The birth of edge cities in China: Measuring the effects of industrial parks policy." *Journal of Urban Economics* 100 80–103.

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*.

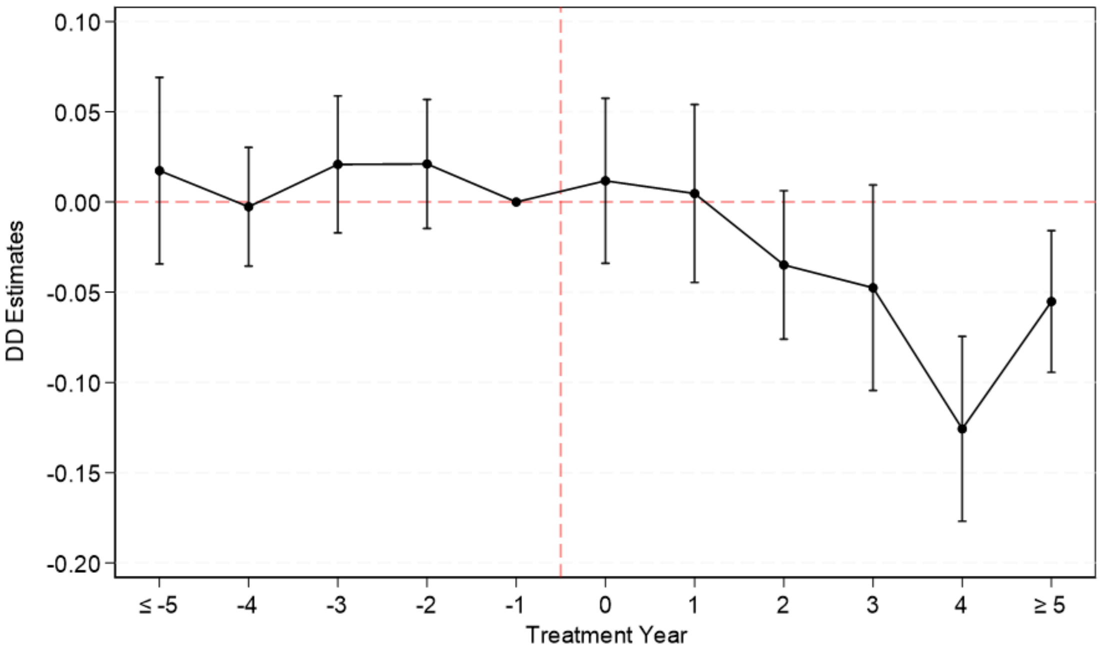
9 Figures and Tables

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



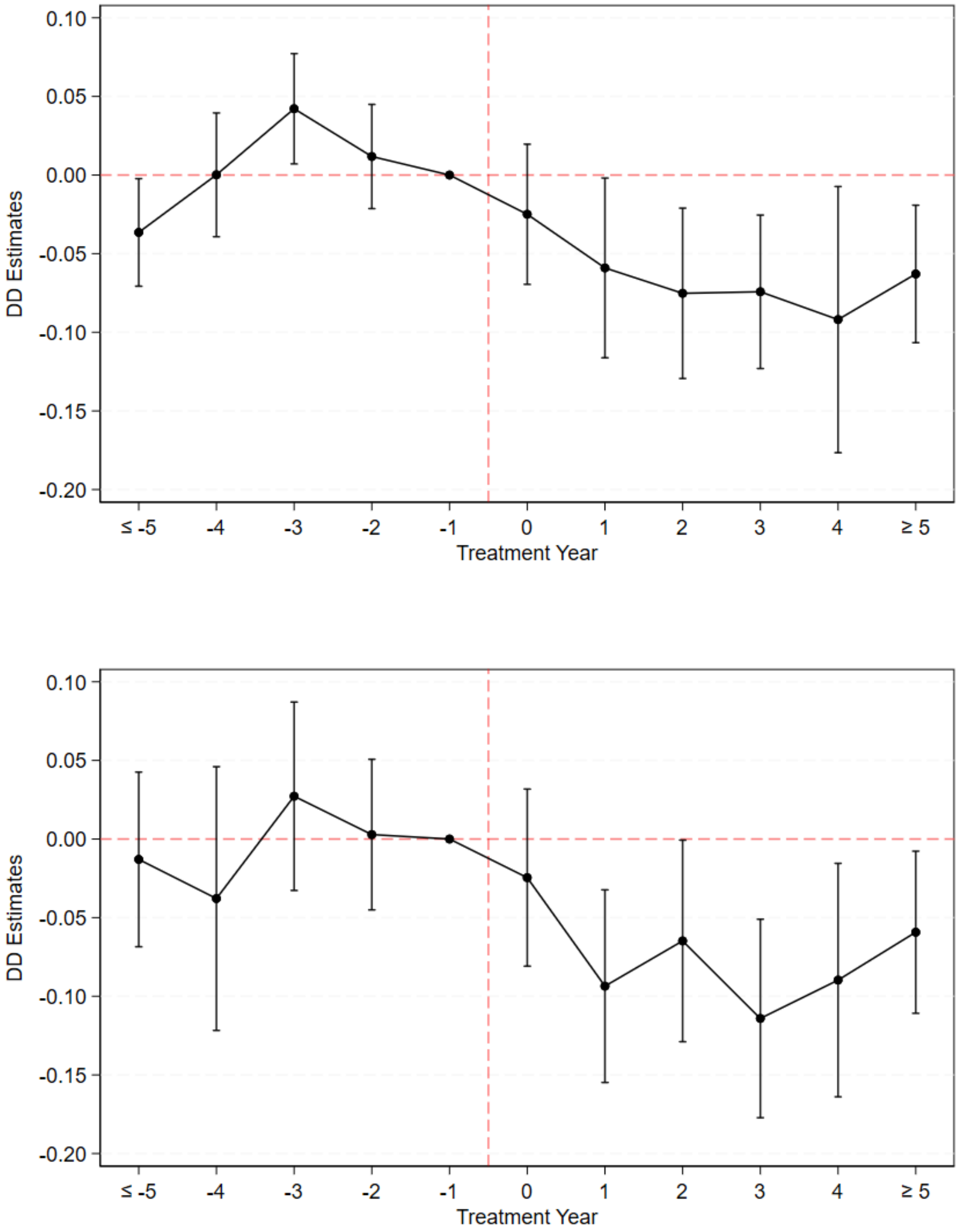
Notes: The figure plots the number of data centers by construction-completion year (top panel) and their geographic distribution across the contiguous United States (bottom panel). Red bars in the top panel and red dots in the bottom panel denote hyperscale data centers, while the gray counterparts represent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

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



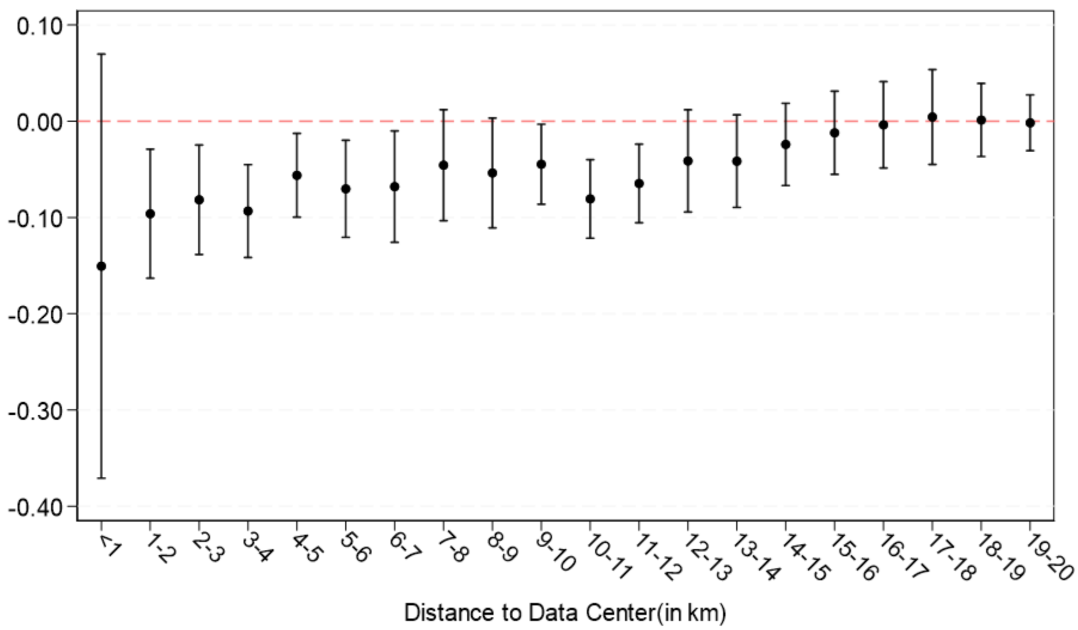
Notes: This figure plots the estimated event-time coefficients from the baseline difference-in-differences model specified in Equation 2. The horizontal axis denotes years relative to the facility opening year (year 0). The coefficient for year -1 is normalized to zero. The sample includes housing transactions within 5 km of hyperscale data centers (treatment group) and transactions 20–50 km from the same facilities (comparison group). Dots represent point estimates $\hat{\beta}_\ell$ and the lines show 95% confidence intervals.

Figure 3: Event Study Using Later-Treated as Control



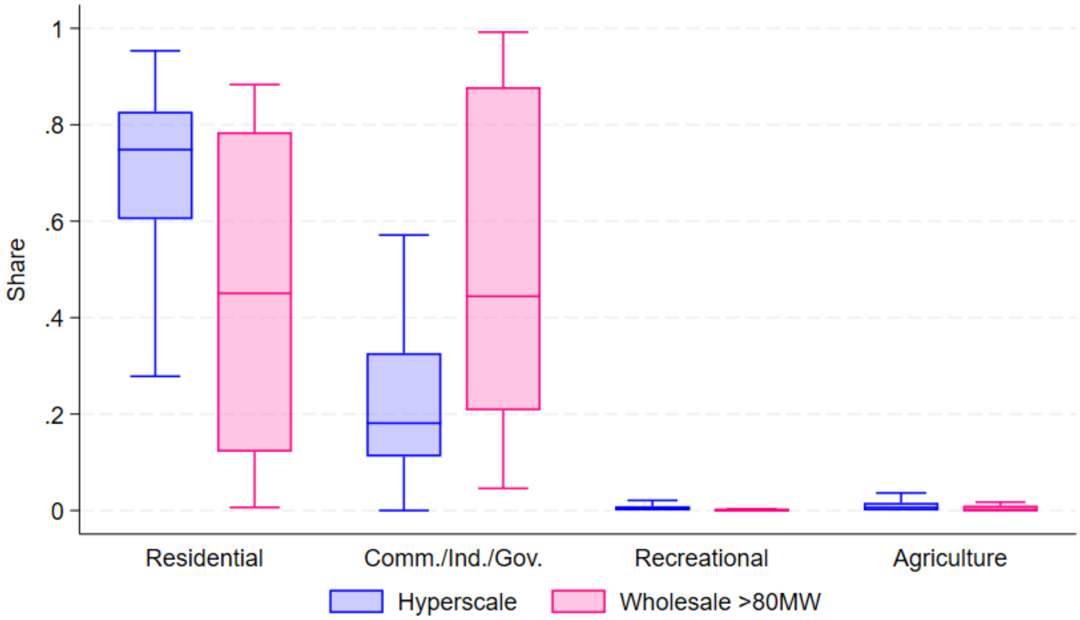
Notes: This figure plots the estimated event-time coefficients from Equation 2, replacing the outer-ring comparison group with the near ring ($\leq 5km$) around facilities that open later. The horizontal axis denotes years relative to the built year of the hyperscale data center (year 0 = year built). The point at -1 (the omitted category) is normalized to zero. The sample consists of housing transactions within 5 km of hyperscale data centers, which serve as the treated group. The upper panel uses all transactions within 5 km of later-opening hyperscale facilities in the same county as the comparison group. The lower panel instead uses transactions within 5 km of later-opening hyperscale facilities in the same county that are matched to treated sites at the census-tract level using XGBoost-predicted siting probabilities, with each treated tract matched to its five nearest later-treated tracts in predicted-probability space. Dots represent point estimates $\hat{\beta}_\ell$ and the lines show 95% confidence intervals.

Figure 4: Housing Price Impact by Distance from Hyperscale Data Center



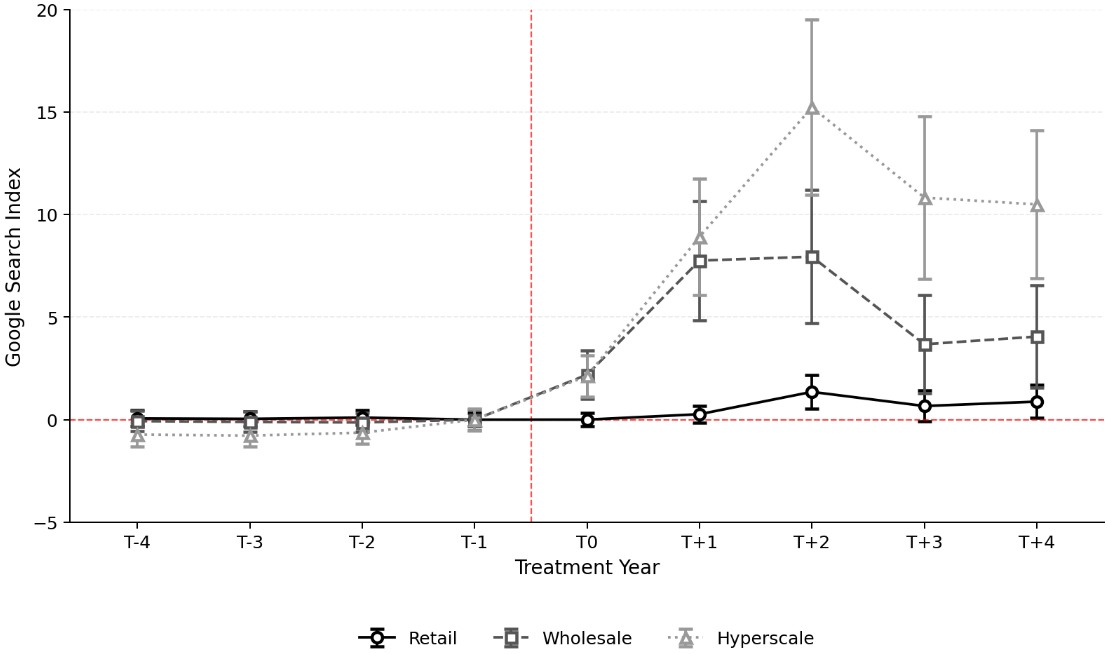
Notes: This figure plots estimated treatment effects on log housing price obtained from 20 separate regressions based on the baseline difference-in-differences specification in Equation 2. Each equation defines the treated sample as transactions within a specific distance band from the first-built hyperscale data center (0-1km, 1-2km,..., 19-20km), while the control group includes transactions located 20–50km from the same facility. Dots represent point estimates $\hat{\beta}_\ell$ and the lines show 95% confidence intervals.

Figure 5: Baseline Surrounding Land Use by Data Center Type



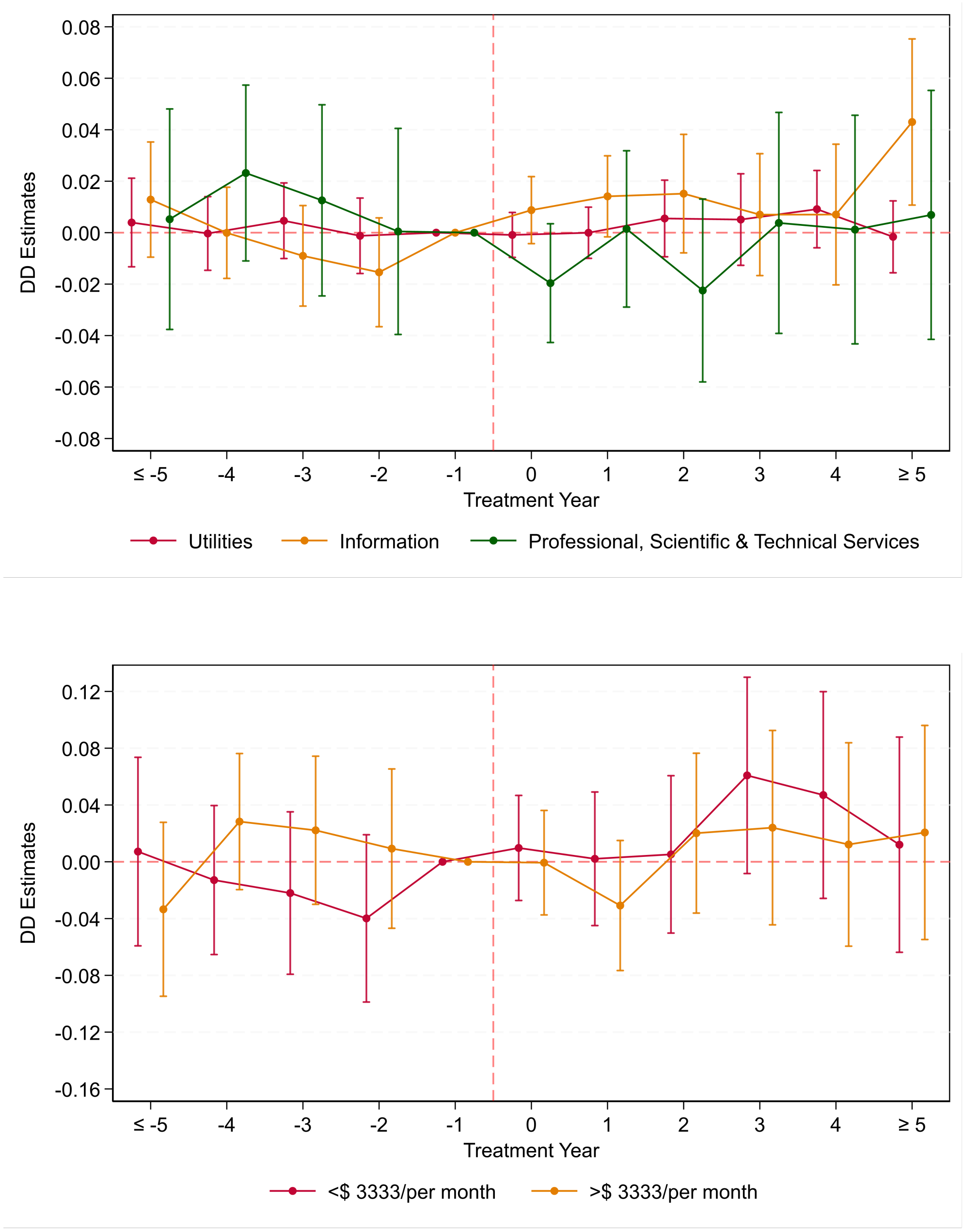
Notes: The figure plots baseline land-use shares within 5 km of data centers, comparing hyperscale facilities with wholesale facilities with Total Utility Power above 80 MW. Land-use data are drawn from Ahn et al., 2024. For each data center, shares are computed using the most recent land-use layer preceding the facility's opening and measure the fraction of nearby land area classified as residential, commercial/industrial/government, recreational, or agricultural.

Figure 6: Google Search Intensity by Data Center Type



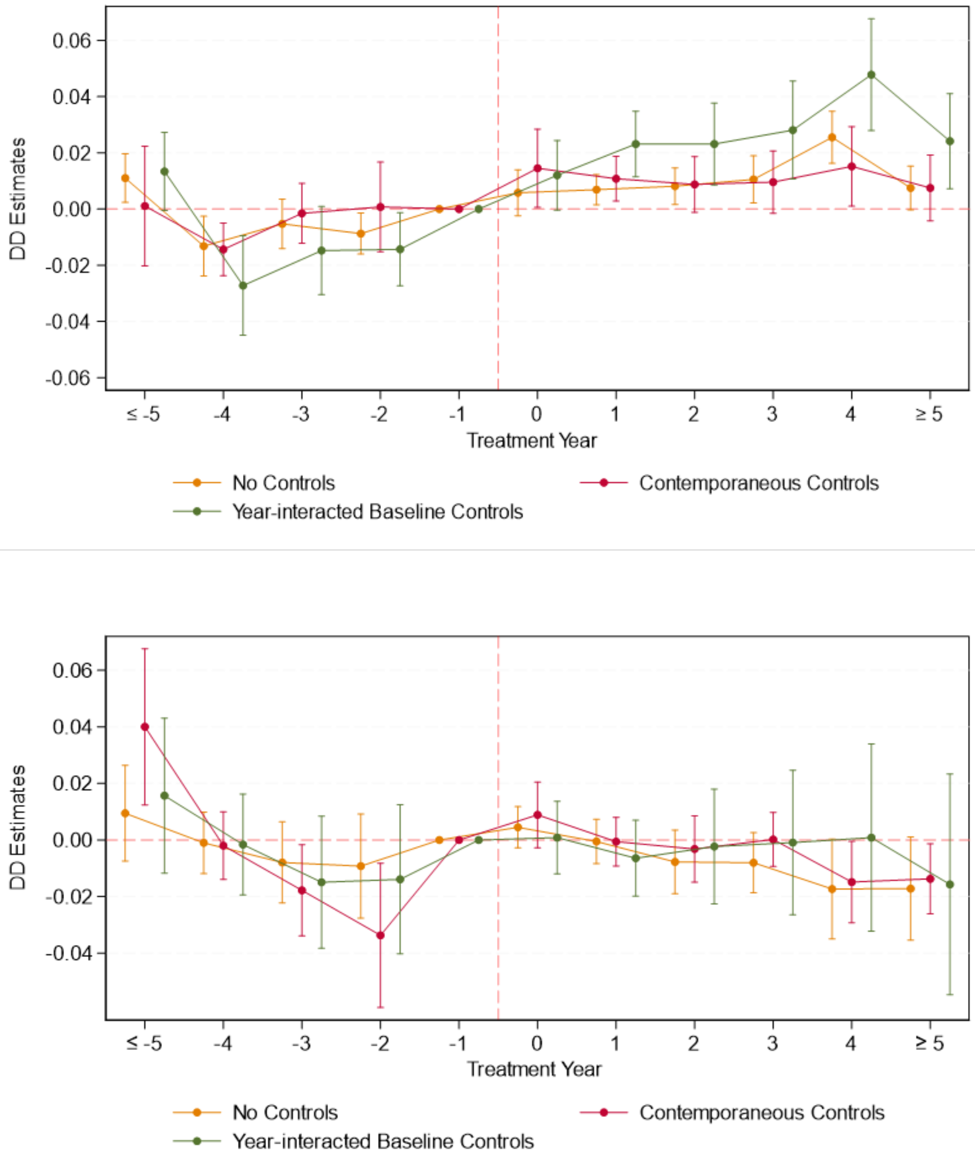
Notes: This figure plots the event study estimates of Google Search Index around data center openings by facility type. The outcome variable is the Google Search Index for the search term ”data center” at the state level, sourced from Google Trends. The index is normalized to zero in the year prior to opening (T-1). Each point represents the average index value across facilities of that type in the corresponding event year, with vertical bars indicating 95% confidence intervals. The dashed vertical red line marks the year of opening (between T-1 and T0). The dashed horizontal red line denotes the baseline value of zero.

Figure 7: Labor Market Impact of Hyperscale Data Center



Notes: This figure plots event-time coefficients for labor-market outcomes in natural logarithms. The upper panel reports estimates for employment in relevant sectors, and the lower panel reports estimates for employment in relevant earnings bins. 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 8: Utility 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: Data Center Types and Baseline Characteristics

Panel A: Distribution and Definitions by Data Center Type

Type	**Number (Percent)**	**Description**
Retail	540 (51.4%)	Multi-tenant facilities that rent space in small units (e.g., racks, cabinets, or cages) and typically bundle managed services such as security, power backup, HVAC, and remote-hands support.
Wholesale	194 (18.5%)	Facilities that lease much larger blocks of space (e.g., pods/rooms) to a small number of tenants—often enterprises or colocation providers.
Hyperscale	162 (15.4%)	Capacity built, owned and used by major hyperscalers including Amazon, Apple, Google, Meta, and Microsoft.
Cryptomining	112 (10.7%)	Specialized facilities primarily used for crypto mining, with highly electricity-intensive and typically more flexible (price-responsive) operations than conventional cloud/enterprise data centers.
Others	43 (4.0%)	A residual category including powered-shell sites (basic building with power/connectivity but unfinished interior), telecom/network facilities, and smaller hosting or cloud-oriented data centers.

Panel B: Facility Characteristics and Baseline Characteristics by Data Center Type

	Non-Hyperscale DC			**Hyperscale DC**			
	(1) Mean	(2) Min	(3) Max	(4) Mean	(5) Min	(6) Max	(1) – (4) Diff
DC Characteristics							
Year Built	2014.324	2001	2025	2018.858	2003	2025	-4.534***
Gross Total Space ('000 sqft)	122.498	0.250	1875.000	301.554	10.000	1335.600	-179.056***
Total Operational Space ('000 sqft)	56.817	0.200	1250.000	197.229	6.700	890.400	-140.412***
Total UPS Power (MW)	13.402	0.020	500.000	38.583	1.340	240.000	-25.181***
Total Facility Power (MW)	17.200	0.030	550.000	48.593	1.742	264.000	-31.393***
Total Utility Power (MW)	20.480	0.030	759.000	56.704	2.010	264.000	-36.224***
Total # Racks	1289.696	6.000	15062.000	5709.722	191.000	25440.000	-4420.026***
Watt/sqft	194.394	48.000	2200.000	198.662	70.000	480.000	-4.268
KW/rack	5.386	1.300	23.300	6.987	2.500	16.800	-1.601***
Demographics (1 year before treatment)							
Total Population (# per sqkm)	1.141	0.000	21.934	0.291	0.002	1.924	0.849***
Total Housing Units (# per sqkm)	0.546	0.000	9.089	0.110	0.001	0.709	0.437***
% of White	0.653	0.002	0.991	0.722	0.160	0.980	-0.068**

% of Black	0.151	0.000	0.957	0.096	0.000	0.724	0.055**
% of Hispanic	0.180	0.000	0.965	0.150	0.003	0.738	0.030
% Less than 9th Grade	0.051	0.000	0.365	0.052	0.000	0.284	-0.002
% of Unemployed	0.071	0.000	0.406	0.067	0.000	0.208	0.003
% of Owner Occupied	0.474	0.000	0.982	0.661	0.008	0.960	-0.187***
Per Capita Income ($'000)	33.342	5.192	125.748	28.104	14.316	60.740	5.238*
Household Income ($'000)	62.866	11.178	221.138	65.031	33.282	172.993	-2.165
Land Use Characteristics within 5km (pre-treatment 5-year interval)							
% Residential	0.484	0.000	0.958	0.65	0.000	0.953	-0.168***
% Commercial/Industrial/Governmental	0.445	0.000	1.000	0.288	0.000	1.000	0.157***
% Recreational	0.016	0.000	0.158	0.009	0.000	0.056	0.006***
% Agricultural	0.011	0.000	0.333	0.017	0.000	0.222	-0.006*
Workplace Characteristics (1 year before treatment)							
Total # Jobs	629.319	1.000	7767.000	343.600	2.000	2417.000	285.719**
Total # Jobs in Industrial Sector	162.142	0.000	3094.000	147.133	0.000	2184.000	15.009
Total # Jobs in Logistics Sector	70.165	0.000	2129.000	73.078	0.000	882.000	-2.912
Total # Jobs in Business Services	242.866	0.000	5302.000	72.556	0.000	2262.000	170.311**
Total # Jobs in Consumer Services	154.145	0.000	6360.000	50.833	0.000	505.000	103.312*

Notes: Panel A reports number distribution and descriptions by data center type. Panel B reports summary statistics for data center characteristics and baseline local-area attributes, separately for non-hyperscale and hyperscale facilities. Non-hyperscale facilities include all facilities other than hyperscale sites, including Retail, Wholesale, Cryptomining, and Others. Demographic variables are measured at the census-tract level using ACS 5-year estimates; all are measured one year prior to the treatment year (the facility's year built). Workplace characteristics are based on LEHD LODES Workplace Area Characteristics (WAC) at the census-block level and are measured one year prior to the year built. DC characteristics are measured at the facility level: Total UPS power (MW) is installed usable UPS capacity supporting IT load at the rack; Total facility power (MW) is total designed electrical capacity (IT plus supporting infrastructure); and Total utility power (MW) is upstream utility power required, computed as Total UPS power × PUE, where PUE (power usage effectiveness) scales IT load to total facility load including non-IT overhead (e.g., cooling and power conversion). Statistical significance of mean differences follows: *p<0.10, **p<0.05, and ***p<0.01

Table 2: Summary Statistics of Property Transaction Characteristics by Distance to Data Centers

	Mean Values of All Data Center		**Mean Values of Hyperscale Data Center**	
	$\leq$ 5 km	20–50 km	$\leq$ 5 km	20–50 km
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: This table reports transaction-level mean characteristics for residential properties within 5 km of a data center and in the 20–50 km, shown separately for (i) all data center openings and (ii) hyperscale openings. Sale price is expressed in thousands of 2010 dollars. "Local buyer/seller" indicates that the buyer/seller is located in the same county as the transacted property, and "corporate buyer/seller" indicates that the buyer/seller is a corporate entity. Property and transaction characteristics are obtained from the Corelogic database and are measured at the time of transaction. Neighborhood demographics are measured at the census-tract level using ACS 5-year estimates (matched to the tract containing the property) and correspond to the period prior to the transaction.

Table 3: 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						
$\leq$ 5 km × 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						
$\leq$ 5 km × 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 × Distance FE	Yes	Yes	Yes	Yes	Yes	Yes
DC × Census Tract FE	Yes	Yes	Yes	Yes	Yes	Yes
DC × Time FE	Yes	Yes	Yes	Yes	No	No
DC × State × Time FE	No	No	No	No	Yes	No
DC × County × Time FE	No	No	No	No	No	Yes

Notes: This table presents ordinary least squares estimates from estimating Equation (2). The dependent variable is the natural logarithm of the residential transaction price adjusted to 2010 dollars. The explanatory variable of interest is $\leq$ 5 km × Post, which equals one for transactions of properties located within 5 km of the first-built data center (Panel A) or the first-built hyperscale data center (Panel B) occurring after the facility is built. The sample is restricted to residential transactions within 50 km of the relevant facility, and observations in the 5–20 km buffer ring are excluded. Transaction controls include indicators for cash purchases, resale transactions, short sales, foreclosure REO transactions, corporate buyers, local buyers, corporate sellers, local sellers, and whether the seller previously purchased the unit using cash, through a short sale, or through a foreclosure REO transactions. Property controls include an indicator for mobile-home transactions, property type (e.g., single-family, duplex), and indicators for views and key structural attributes (e.g., fuel type, water/utility/sewer type, electrical wiring, roof type and cover, building quality and improvement condition, heating type, floor type, basement, pool), as well as percentiles of number bedrooms, bathrooms, property age, land area, living area, number of stories, and parking. All categorical covariates include an additional missing category. Distance fixed effects control for property–facility distance using 100 bins of width 0.1 km. Standard errors are clustered at the census-tract level and reported in parentheses. *, **, and *** denote significance at the 10%, 5%, and 1% levels, respectively.

Table 4: Heterogeneity by Capacity

	Log Housing Sale Price			
	(1)	**(2)**	**(3)**	**(4)**
$\leq$5 km × Post	-0.240^{***} (0.032)	-0.088^{***} (0.030)	-0.092^{***} (0.026)	-0.064^{**} (0.026)
Transaction Controls	Yes	Yes	Yes	Yes
Property Controls	Yes	Yes	Yes	Yes
DC × Distance-to-DC FE	Yes	Yes	Yes	Yes
DC × Census Tract FE	Yes	Yes	Yes	Yes
DC × County × Time FE	Yes	Yes	Yes	Yes
Adj. R^2	0.604	0.603	0.629	0.655
Observations	451,175	547,221	1,237,464	2,178,044
Total Utility Power (MW)	$\geq$100	$\geq$80	$\geq$60	$\geq$40

Notes: This table presents ordinary least squares estimates from estimating Equation (2). The dependent variable is the natural logarithm of the residential transaction price adjusted to 2010 dollars. The explanatory variable of interest is $\leq$ 5 km × Post, which equals one for transactions of properties located within 5 km of the first-built hyperscale data center (defined using Total Uninterruptible Power Supply) occurring after the facility is built. Columns (1)–(4) define the treatment sample to first-built facilities whose total UPS capacity is at least 100 MW, 80 MW, 60 MW, and 40 MW, respectively. The sample is restricted to residential transactions within 50 km of the relevant facility, and observations in the 5–20 km buffer ring are excluded. Transaction controls include indicators for cash purchases, resale transactions, short sales, foreclosure REO transactions, corporate buyers, local buyers, corporate sellers, local sellers, and whether the seller previously purchased the unit using cash, through a short sale, or through a foreclosure REO transactions. Property controls include an indicator for mobile-home transactions, property type (e.g., single-family, duplex), and indicators for views and key structural attributes (e.g., fuel type, water/utility/sewer type, electrical wiring, roof type and cover, building quality and improvement condition, heating type, floor type, basement, pool), as well as percentiles of number bedrooms, bathrooms, property age, land area, living area, number of stories, and parking. All categorical covariates include an additional missing category. Distance fixed effects control for property–facility distance using 100 bins of width 0.1 km. Standard errors are clustered at the census-tract level and reported in parentheses. *, **, and *** denote significance at the 10%, 5%, and 1% levels, respectively.

Table 5: Subsample Analysis by Zoning Jurisdiction

	Log Housing Sale Price					
	(1)	**(2)**	**(3)**	**(4)**	**(5)**	**(6)**
≤5 km × Post	0.003	−0.090***	−0.001	−0.146***	−0.019	−0.078***
(0.031)	(0.026)	(0.030)	(0.024)	(0.029)	(0.021)	
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	0.616	0.756	0.660	0.720	0.660	0.662
Observations	1,052,440	180,798	3,319,532	338,573	3,497,591	3,604,160
Treatment Sample	Different NZA	Same NZA	Different NZA/ County	Same NZA/ County	Different NZA/ County	Same NZA/ County
Control Sample	Different NZA	Same NZA	Different NZA/ County	Same NZA/ County	All	All

Notes: This table presents ordinary least squares estimates from estimating Equation (2). The dependent variable is the natural logarithm of the residential transaction price adjusted to 2010 dollars. The explanatory variable of interest is ≤ 5 km × Post, which equals one for transactions of properties located within 5 km of the first-built hyperscale data center occurring after the facility is built. The sample is restricted to residential transactions within 50 km of the relevant facility, and observations in the 5–20 km buffer ring are excluded. Subsamples are defined by whether housing transactions and the data center fall in different zoning jurisdictions (Column 1) or the same zoning jurisdiction (Column 2), using National Zoning Atlas (NZA) boundaries. When NZA zoning boundaries are unavailable, county boundaries are used instead; Columns 3 and 4 apply this fallback rule. Columns 5 and 6 use the same classification as Columns 3 and 4 but retain the full control sample regardless of jurisdictional similarity. Transaction controls include indicators for cash purchases, resale transactions, short sales, foreclosure REO transactions, corporate buyers, local buyers, corporate sellers, local sellers, and whether the seller previously purchased the unit using cash, through a short sale, or through a foreclosure REO transactions. Property controls include an indicator for mobile-home transactions, property type (e.g., single-family, duplex), and indicators for views and key structural attributes (e.g., fuel type, water/utility/sewer type, electrical wiring, roof type and cover, building quality and improvement condition, heating type, floor type, basement, pool), as well as percentiles of number bedrooms, bathrooms, property age, land area, living area, number of stories, and parking. All categorical covariates include an additional missing category. Distance fixed effects control for property–facility distance using 100 bins of width 0.1 km. Standard errors are clustered at the census-tract level and reported in parentheses. *, **, and *** denote significance at the 10%, 5%, and 1% levels, respectively.

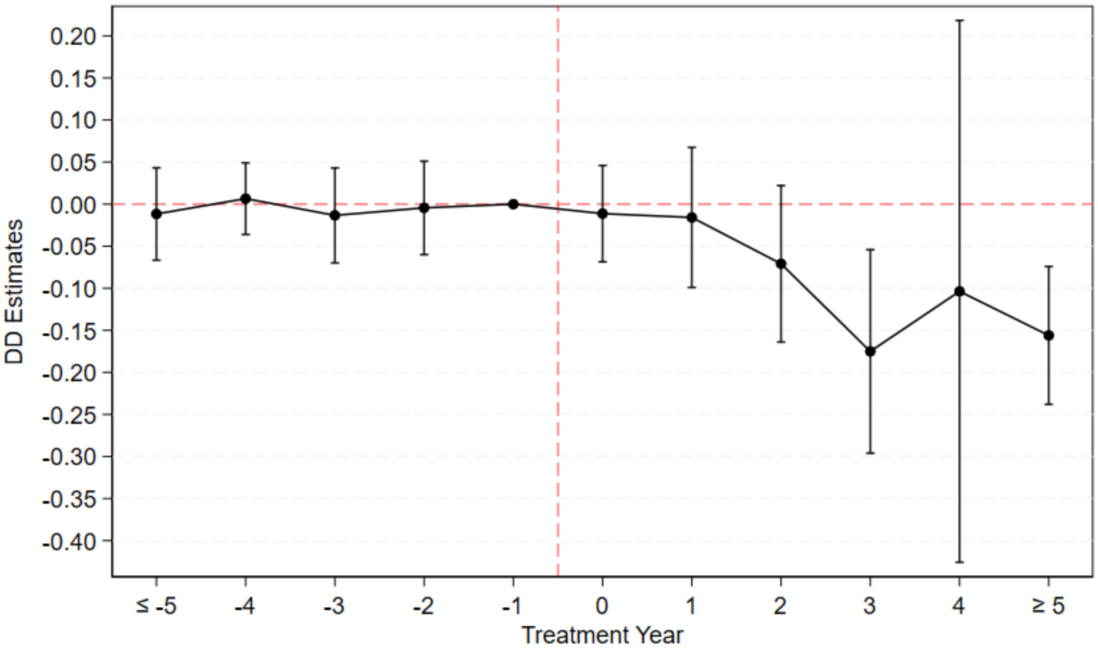
Table 6: Big Tech Ownership and Housing Price Effects

	Log Housing Sale Price		
	(1)	(2)	(3)
$\leq$ 5 km × Post	0.003	0.003	0.003
	(0.005)	(0.005)	(0.005)
$\leq$ 5 km × Post × 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 × Distance FE	Yes	Yes	Yes
DC × Census Tract FE	Yes	Yes	Yes
DC × County × 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: This table presents ordinary least squares estimates from estimating Equation (2), with additional interaction variable with "Big Tech", which is an indicator for data centers operated by Amazon, Meta, Google, Apple, or Microsoft. The dependent variable is the natural logarithm of the residential transaction price adjusted to 2010 dollars. The explanatory variable of interest is $\leq$ 5 km × Post, which equals one for transactions of properties located within 5 km of the first-built hyperscale data center occurring after the facility is built. The sample is restricted to residential transactions within 50 km of the relevant facility, and observations in the 5–20 km buffer ring are excluded. Transaction controls include indicators for cash purchases, resale transactions, short sales, foreclosure REO transactions, corporate buyers, local buyers, corporate sellers, local sellers, and whether the seller previously purchased the unit using cash, through a short sale, or through a foreclosure REO transactions. Property controls include an indicator for mobile-home transactions, property type (e.g., single-family, duplex), and indicators for views and key structural attributes (e.g., fuel type, water/utility/sewer type, electrical wiring, roof type and cover, building quality and improvement condition, heating type, floor type, basement, pool), as well as percentiles of number bedrooms, bathrooms, property age, land area, living area, number of stories, and parking. All categorical covariates include an additional missing category. Distance fixed effects control for property–facility distance using 100 bins of width 0.1 km. Standard errors are clustered at the census-tract level and reported in parentheses. *, **, and *** denote significance at the 10%, 5%, and 1% levels, respectively.

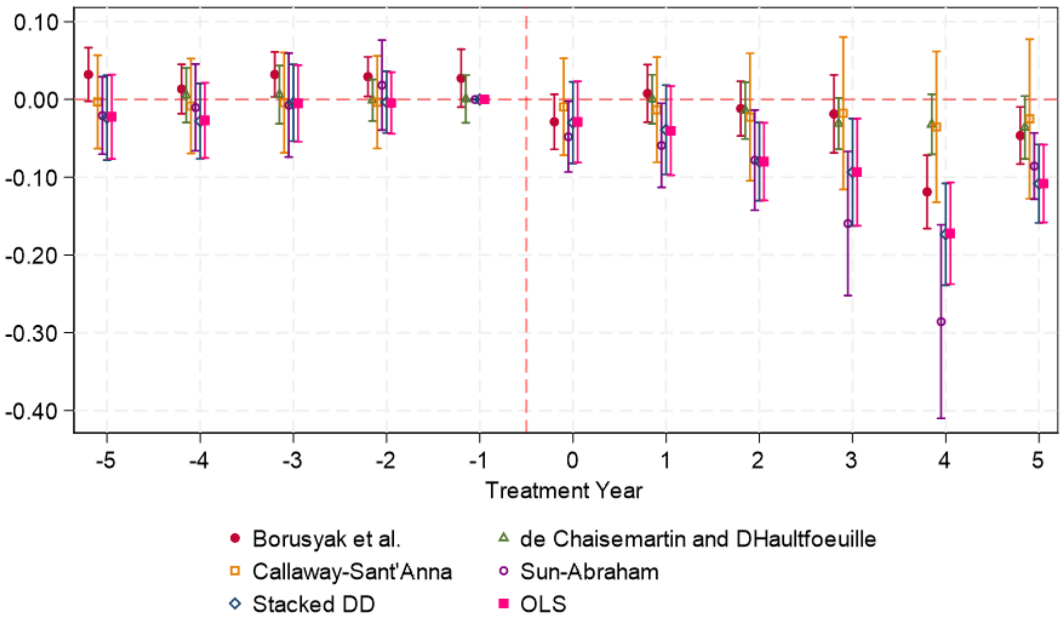
A Appendix Figures and Tables

Figure A-1: Effects of Hyperscale Data Center Openings on Housing Prices—exclude transactions after experiencing subsequent entries



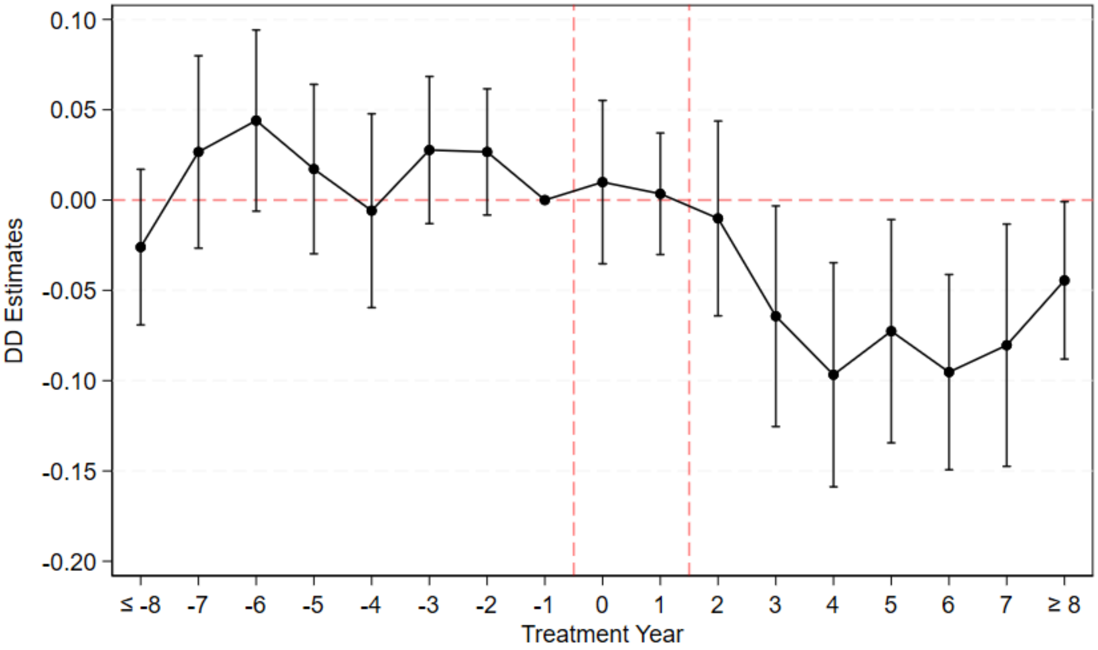
Notes: This figure plots the estimated event-time coefficients from the baseline difference-in-differences model specified in Equation 2. The horizontal axis denotes years relative to the facility opening year (year 0). The coefficient for year -1 is normalized to zero. The sample includes housing transactions within 5 km of hyperscale data centers (treatment group) and transactions 20–50 km from the same facilities (comparison group). The treated group excludes transactions occurring in the year of, and after, the opening of any subsequent data centers. Dots represent point estimates $\hat{\beta}_\ell$ and the lines show 95% confidence intervals.

Figure A-2: Staggered DiD Results of Log Housing Prices Around Hyperscale Data Center Openings



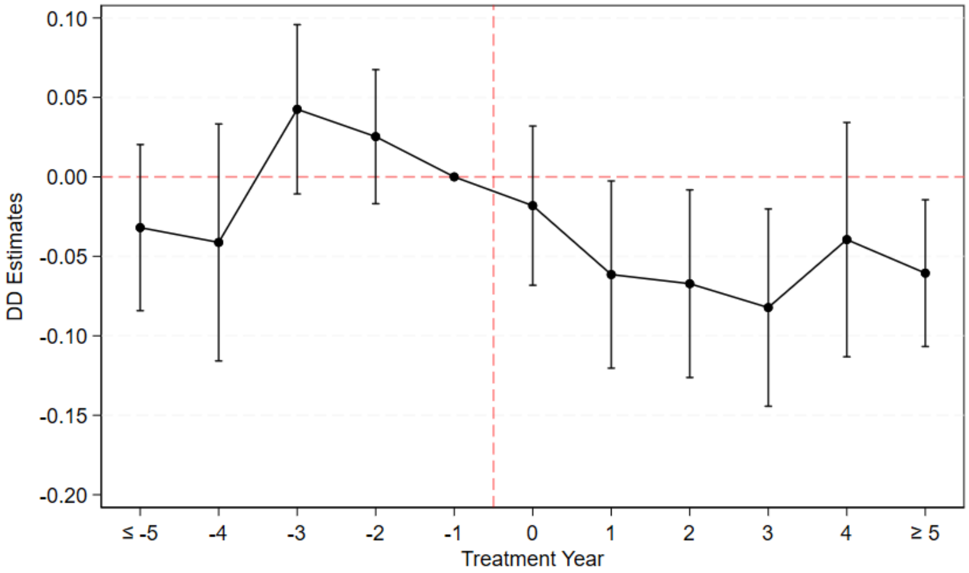
Notes: This figure plots the estimated event-time coefficients from alternative staggered difference-in-differences estimators. The horizontal axis denotes years relative to the facility opening year (year 0). The coefficient for year −1 is normalized to zero. The sample includes housing transactions within 5 km of hyperscale data centers (treatment group) and transactions 20–50 km from the same facilities (comparison group). Dots represent point estimates $\hat{\beta}_\ell$ and the lines show 95% confidence intervals.

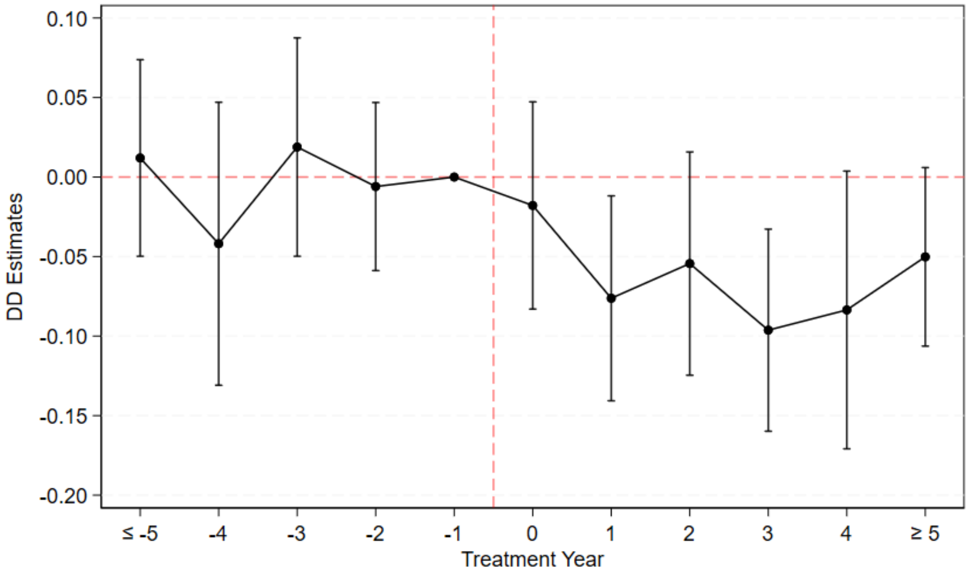
Figure A-3: Impacts of Hyperscale Data Center Announcements/Constructions/Initial Engagements on Housing Prices



Notes: This figure plots the estimated event-time coefficients from the baseline difference-in-differences model specified in Equation 2. The horizontal axis denotes years relative to the earliest of three project milestones for the hyperscale data center—the announcement year, construction-start year, or initial public-engagement year (year 0 = earliest milestone year). The coefficient for year -1 is normalized to zero. The sample includes housing transactions within 5 km of hyperscale data centers (treatment group) and transactions 20–50 km from the same facilities (comparison group). Dots represent point estimates $\hat{\beta}_\ell$ and the lines show 95% confidence intervals.

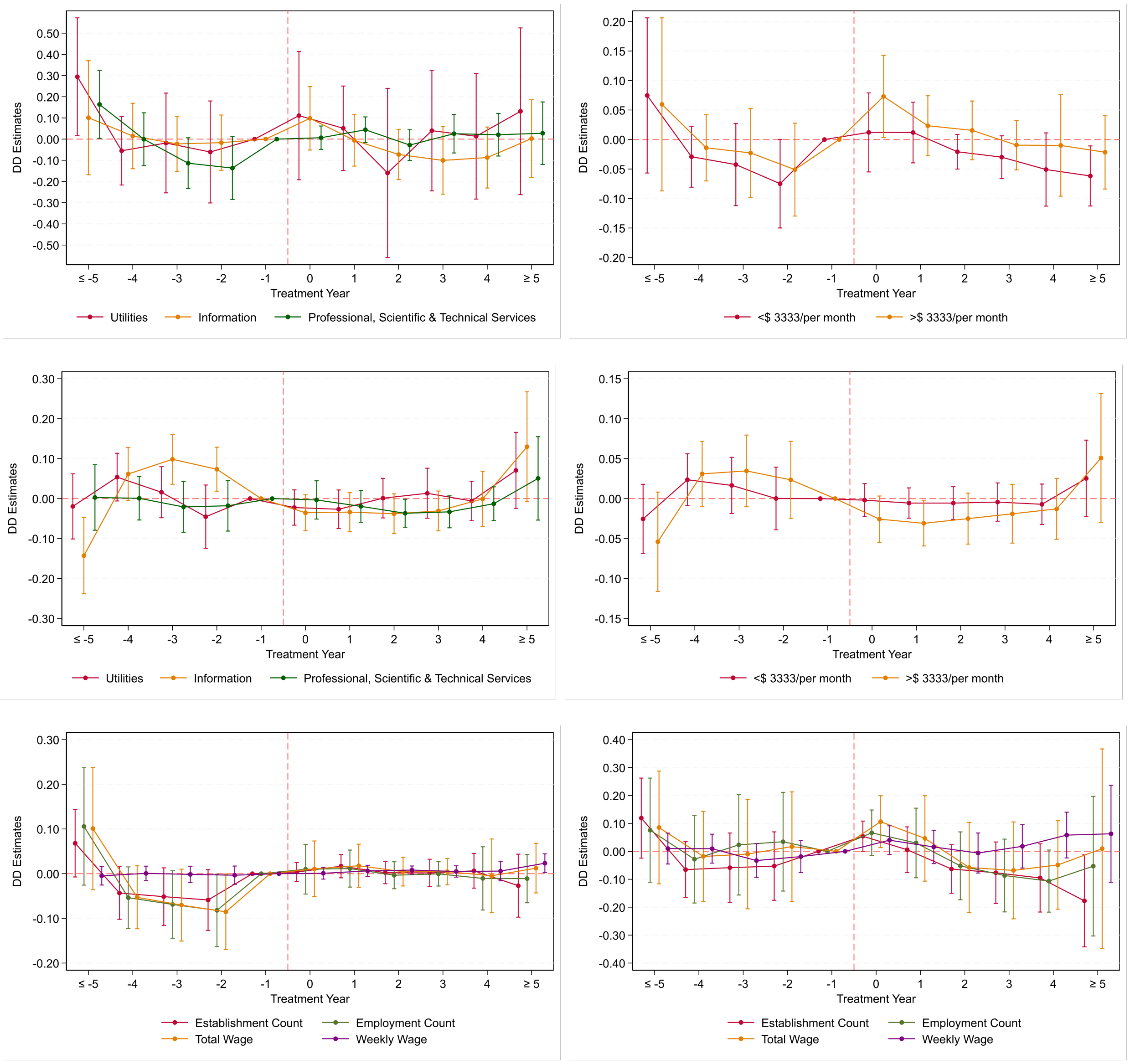
Figure A-4: Event Study Using XGBoost-matched Later-Treated as Control





Notes: This figure plots the estimated event-time coefficients from Equation 2, replacing the outer-ring comparison group with the near ring ($\leq 5km$) around facilities that open later. The horizontal axis denotes years relative to the built year of the hyperscale data center (year 0 = year built). The point at -1 (the omitted category) is normalized to zero. The sample consists of housing transactions within 5 km of hyperscale data centers, which serve as the treated group. Both panels use transactions within 5 km of later-opening hyperscale facilities in the same county as the comparison group, matched at the census-tract level using XGBoost-predicted siting probabilities. The upper panel uses ten nearest-neighbor matches for each treated tract, while the lower panel uses three nearest-neighbor matches. Dots represent point estimates $\hat{\beta}_\ell$ and the lines show 95% confidence intervals.

Figure A-5: Labor Market Impact of Hyperscale Data Center—at County and Commuting Zone Level



Notes: This figure plots event-time coefficients for labor-market outcomes in natural logarithms. Labor market outcomes in the upper and middle panels are obtained from LODES, and are aggregated at the county or commuting zone level; labor market outcomes in the bottom panel are at the county level and are obtained from QCEW. For the upper and middle panels, the left figures report estimates for employment in relevant sectors, and the right figures reports estimates for employment in relevant earnings bins. For the bottom panel, the left figure reports labor-market outcomes for all sectors, and the right figure reports labor-market outcomes for the Information sector only. Treated units are counties (commuting zone) hosting a hyperscale data center, and the comparison group consists of later-treated counties (commuting zone) within the same state. 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: Data Center Types and Surrounding Building Structures

	Non-Hyperscale DC			**Hyperscale DC**			
	(1) Mean	(2) Min	(3) Max	(4) Mean	(5) Min	(6) Max	(1) – (4) Diff
Total Size ('000 Sqft) of Building Structures by Occupancy Classification							
Assembly	82.797	0.000	1794.035	4.547	0.000	152.101	78.250***
Commercial	1153.034	0.000	7517.541	464.390	0.000	3340.147	688.644***
Education	193.110	0.000	3208.816	25.099	0.000	633.552	168.011***
Government	142.073	0.000	4319.909	39.291	0.000	1762.917	102.782***
Industrial	42.484	0.000	2372.299	6.930	0.000	131.211	35.554**
Industrial – High Tech	25.649	0.000	1426.058	19.827	0.000	1489.400	5.822
Industrial – Light	928.469	0.000	12028.758	290.524	0.000	3906.754	637.945***
Total Number of Building Structures by Occupancy Classification)							
Assembly	5.285	0.000	87.000	0.710	0.000	26.000	4.575***
Commercial	82.813	0.000	533.000	22.914	0.000	179.000	59.900***
Education	9.829	0.000	152.000	1.333	0.000	23.000	8.496***
Government	8.712	0.000	168.000	2.975	0.000	49.000	5.737***
Industrial	1.645	0.000	48.000	0.691	0.000	16.000	0.953*
Industrial – High Tech	0.633	0.000	38.000	0.611	0.000	25.000	0.022
Industrial – Light	28.377	0.000	340.000	7.821	0.000	66.000	20.556***

Notes: This table reports summary statistics for building structures within 1 km of data centers, separately for non-hyperscale and hyperscale facilities. Non-hyperscale facilities include all facilities other than hyperscale sites, including Retail, Wholesale, Cryptomining, and Others. Building footprint and occupancy classification data are drawn from FEMA's USA Structures Database, which covers all structures larger than 450 square feet. Assembly includes community centers, convention centers, indoor arenas, religious buildings, and stadiums; Commercial includes retail and wholesale trade, professional and technical services, entertainment and recreation, personal and repair services, medical facilities, and banks; Education includes Pre-K–12 schools, colleges, universities, and other educational buildings; Government includes general services, emergency response, and non-civilian structures; Industrial includes heavy industry, metal and mineral processing, food, drug and chemical production, and construction. Statistical significance of mean differences follows: *p<0.10, **p<0.05, and ***p<0.01.

Table A-2: Effects of Hyperscale Data Centers on House Prices—Alternative Control Rings

	Log Housing Sale Price			
	(1)	**(2)**	**(3)**	**(4)**
Treated × Post	-0.034^{***}	-0.056^{**}	-0.053^{***}	-0.089^{***}
	(0.012)	(0.013)	(0.016)	(0.022)
Transaction Controls	Yes	Yes	Yes	Yes
Property Controls	Yes	Yes	Yes	Yes
DC × Census Tract FE	Yes	Yes	Yes	Yes
DC × County × Time FE	Yes	Yes	Yes	Yes
DC × Distance-to-DC FE	Yes	Yes	Yes	Yes
Distance-to-DC FE	No	No	No	No
Adj. R^2	0.686	0.667	0.676	0.667
Observations	27,834,778	6,013,718	1,568,211	3,119,525
Control Group	5–50 km	15–50 km	15–30 km	30–50km

Notes: This table presents ordinary least squares estimates from estimating Equation (2). The dependent variable is the natural logarithm of the residential transaction price adjusted to 2010 dollars. The explanatory variable of interest is $\leq$ 5 km × Post, which equals one for transactions of properties located within 5 km of the first-built hyperscale data center occurring after the facility is built. Transaction controls include indicators for cash purchases, resale transactions, short sales, foreclosure REO transactions, corporate buyers, local buyers, corporate sellers, local sellers, and whether the seller previously purchased the unit using cash, through a short sale, or through a foreclosure REO transactions. Property controls include an indicator for mobile-home transactions, property type (e.g., single-family, duplex), and indicators for views and key structural attributes (e.g., fuel type, water/utility/sewer type, electrical wiring, roof type and cover, building quality and improvement condition, heating type, floor type, basement, pool), as well as percentiles of number bedrooms, bathrooms, property age, land area, living area, number of stories, and parking. All categorical covariates include an additional missing category. Distance fixed effects control for property–facility distance using 100 bins of width 0.1 km. The bottom panel outlines the control groups used in each specification across columns. Standard errors are clustered at the census-tract level and reported in parentheses. *, **, and *** denote significance at the 10%, 5%, and 1% levels, respectively.

Table A-3: Effects of Smaller-Scale Data Centers on House Prices

	Log Housing Sale Price				
	(1)	(2)	(3)	(4)	(5)
Treatment × Post	−0.017 (0.015)	0.007 (0.007)	0.002 (0.008)	−0.003 (0.006)	0.000 (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	4–5 km
Control Group			20–50km		

Notes: This table presents ordinary least squares estimates from estimating Equation (2). The dependent variable is the natural logarithm of the residential transaction price adjusted to 2010 dollars. The explanatory variable of interest is ≤ Treatment × Post. Treatment is defined as transactions of properties located within 5 km (Column 1), 1–2km (Column 2), 2–3km (Column 3), 3–4km (Column 4), and 4–5 km (Column 5) of the non-hyperscale data center. The control sample includes residential transactions within 20–50 km of the relevant facility. Transaction controls include indicators for cash purchases, resale transactions, short sales, foreclosure REO transactions, corporate buyers, local buyers, corporate sellers, local sellers, and whether the seller previously purchased the unit using cash, through a short sale, or through a foreclosure REO transactions. Property controls include an indicator for mobile-home transactions, property type (e.g., single-family, duplex), and indicators for views and key structural attributes (e.g., fuel type, water/utility/sewer type, electrical wiring, roof type and cover, building quality and improvement condition, heating type, floor type, basement, pool), as well as percentiles of number bedrooms, bathrooms, property age, land area, living area, number of stories, and parking. All categorical covariates include an additional missing category. Distance fixed effects control for property–facility distance using 100 bins of width 0.1 km. Standard errors are clustered at the census-tract level and reported in parentheses. *, **, and *** denote significance at the 10%, 5%, and 1% levels, respectively.

Table A-4: Heterogeneity by Capacity—Wholesale Data Center

	Log Housing Sale Price			
	(1)	**(2)**	**(3)**	**(4)**
≤5 km × Post	−0.004	−0.000	0.002	0.022
	(0.030)	(0.030)	(0.025)	(0.016)
Transaction Controls	Yes	Yes	Yes	Yes
Property Controls	Yes	Yes	Yes	Yes
DC × Distance-to-DC FE	Yes	Yes	Yes	Yes
DC × Census Tract FE	Yes	Yes	Yes	Yes
DC × County × Time FE	Yes	Yes	Yes	Yes
Adj. R^2	0.647	0.647	0.653	0.655
Observations	284,736	331,353	817,514	2,314,594
Total Utility Power (MW)	≥100	≥80	≥60	≥40

Notes: This table presents ordinary least squares estimates from estimating Equation (2). The dependent variable is the natural logarithm of the residential transaction price adjusted to 2010 dollars. The explanatory variable of interest is ≤ 5 km × Post, which equals one for transactions of properties located within 5 km of the first-built wholesale data center (defined using Total Utility Power) occurring after the facility is built. Columns (1)–(4) define the treatment sample to first-built facilities whose Total Utility Power is at least 100 MW, 80 MW, 60 MW, and 40 MW, respectively. The sample is restricted to residential transactions within 50 km of the relevant facility, and observations in the 5–20 km buffer ring are excluded. Transaction controls include indicators for cash purchases, resale transactions, short sales, foreclosure REO transactions, corporate buyers, local buyers, corporate sellers, local sellers, and whether the seller previously purchased the unit using cash, through a short sale, or through a foreclosure REO transactions. Property controls include an indicator for mobile-home transactions, property type (e.g., single-family, duplex), and indicators for views and key structural attributes (e.g., fuel type, water/utility/sewer type, electrical wiring, roof type and cover, building quality and improvement condition, heating type, floor type, basement, pool), as well as percentiles of number bedrooms, bathrooms, property age, land area, living area, number of stories, and parking. All categorical covariates include an additional missing category. Distance fixed effects control for property–facility distance using 100 bins of width 0.1 km. Standard errors are clustered at the census-tract level and reported in parentheses. *, **, and *** denote significance at the 10%, 5%, and 1% levels, respectively.

Table A-5: Heterogeneity by Surrounding —Non-Hyperscale Data Center

	Log Housing Sale Price			
	(1)	**(2)**	**(3)**	**(4)**
$\leq$5 km × Post	-0.089^{**} (0.035)	0.026 (0.031)	-0.006 (0.040)	-0.031 (0.028)
Transaction Controls	Yes	Yes	Yes	Yes
Property Controls	Yes	Yes	Yes	Yes
DC × Distance-to-DC FE	Yes	Yes	Yes	Yes
DC × Census Tract FE	Yes	Yes	Yes	Yes
DC × County × Time FE	Yes	Yes	Yes	Yes
Adj. R^2	0.678	0.692	0.636	0.686
Observations	319,545	1,436,318	270,599	720,844
Total Utility Power (MW)	$\geq$80	$\geq$40	$\geq$80	$\geq$40
Fraction of Comm./Ind./Gov. Land Uses	Bottom Quartile		Top Quartile	

Notes: This table presents ordinary least squares estimates from estimating Equation (2). The dependent variable is the natural logarithm of the residential transaction price adjusted to 2010 dollars. The explanatory variable of interest is $\leq$ 5 km × Post, which equals one for transactions of properties located within 5 km of the first-built non-hyperscale data center occurring after the facility is built. Columns (1) and (3) , and Columns (2) and (4) define the treatment sample to first-built facilities whose Total Utility Power is 80 MW and 40 MW, respectively. The sample is restricted to residential transactions within 50 km of the relevant facility, and observations in the 5–20 km buffer ring are excluded. The sample in Columns (1) and (2) (Columns (3) and (4)) is further restricted to facilities whose surrounding commercial/industrial/government land share, measured within 5 km using the most recent pre-treatment land-use layer, lies in the bottom (top) quartile of the distribution distribution. Transaction controls include indicators for cash purchases, resale transactions, short sales, foreclosure REO transactions, corporate buyers, local buyers, corporate sellers, local sellers, and whether the seller previously purchased the unit using cash, through a short sale, or through a foreclosure REO transactions. Property controls include an indicator for mobile-home transactions, property type (e.g., single-family, duplex), and indicators for views and key structural attributes (e.g., fuel type, water/utility/sewer type, electrical wiring, roof type and cover, building quality and improvement condition, heating type, floor type, basement, pool), as well as percentiles of number bedrooms, bathrooms, property age, land area, living area, number of stories, and parking. All categorical covariates include an additional missing category. Distance fixed effects control for property–facility distance using 100 bins of width 0.1 km. Standard errors are clustered at the census-tract level and reported in parentheses. *, **, and *** denote significance at the 10%, 5%, and 1% levels, respectively.

Table A-6: Heterogeneity by County-level Number of Prior Entries and Clusters

	Log Housing Sale Price				
	(1)	**(2)**	**(3)**	**(4)**	**(5)**
≤5 km × Post	-0.107^{***} (0.023)	-0.082^{**} (0.039)	-0.063^{**} (0.025)	-0.176^{***} (0.040)	0.050^{**} (0.024)
Transaction Controls	Yes	Yes	Yes	Yes	Yes
Property Controls	Yes	Yes	Yes	Yes	Yes
DC × Distance-to-DC 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	0.631	0.649	0.670	0.663	0.716
Observations	640,723	328,786	2,692,325	1,032,954	372,097
Sample: No. of Entry	1	2	≥ 3		
Sample: DC Clustering Quartile				1	4

Notes: This table presents ordinary least squares estimates from estimating Equation (2). The dependent variable is the natural logarithm of the residential transaction price adjusted to 2010 dollars. The explanatory variable of interest is ≤ 5 km × Post, which equals one for transactions of properties located within 5 km of the first-built hyperscale data center occurring after the facility is built. The sample is restricted to residential transactions within 50 km of the relevant facility, and observations in the 5–20 km buffer ring are excluded. Columns (1)–(3) further restrict the sample by the hyperscale facility's county-level entry order in a county. Columns (4)–(5) instead stratify counties by county-level count of data center measured as of 2025. Transaction controls include indicators for cash purchases, resale transactions, short sales, foreclosure REO transactions, corporate buyers, local buyers, corporate sellers, local sellers, and whether the seller previously purchased the unit using cash, through a short sale, or through a foreclosure REO transactions. Property controls include an indicator for mobile-home transactions, property type (e.g., single-family, duplex), and indicators for views and key structural attributes (e.g., fuel type, water/utility/sewer type, electrical wiring, roof type and cover, building quality and improvement condition, heating type, floor type, basement, pool), as well as percentiles of number bedrooms, bathrooms, property age, land area, living area, number of stories, and parking. All categorical covariates include an additional missing category. Distance fixed effects control for property–facility distance using 100 bins of width 0.1 km. Standard errors are clustered at the census-tract level and reported in parentheses. *, **, and *** denote significance at the 10%, 5%, and 1% levels, respectively.

Table A-7: Water-Service-Area Controls and Housing Price Effects

	Log Housing Sale Price			
	(1)	(2)	(3)	(4)
$\leq$ 5 km × 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 × Distance FE	Yes	Yes	Yes	Yes
DC × Census Tract FE	Yes	Yes	Yes	Yes
DC × Service Boundary × Time FE	Yes	No	Yes	No
DC × County × Service Boundary × 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: This table presents ordinary least squares estimates from estimating Equation (2). The dependent variable is the natural logarithm of the residential transaction price adjusted to 2010 dollars. The explanatory variable of interest is $\leq$ 5 km × Post, which equals one for transactions of properties located within 5 km of the first-built hyperscale data center occurring after the facility is built. Transaction data are further merged with water-service boundary data from SimpleLab TEMM, with each transaction assigned to a boundary. Transactions that cannot be matched to any boundary are excluded in Columns (1) and (2), and are instead coded as a "missing boundary" category and retained in Columns (3) and (4). The sample is restricted to residential transactions within 50 km of the relevant facility, and observations in the 5–20 km buffer ring are excluded. Transaction controls include indicators for cash purchases, resale transactions, short sales, foreclosure REO transactions, corporate buyers, local buyers, corporate sellers, local sellers, and whether the seller previously purchased the unit using cash, through a short sale, or through a foreclosure REO transactions. Property controls include an indicator for mobile-home transactions, property type (e.g., single-family, duplex), and indicators for views and key structural attributes (e.g., fuel type, water/utility/sewer type, electrical wiring, roof type and cover, building quality and improvement condition, heating type, floor type, basement, pool), as well as percentiles of number bedrooms, bathrooms, property age, land area, living area, number of stories, and parking. All categorical covariates include an additional missing category. Distance fixed effects control for property–facility distance using 100 bins of width 0.1 km. Standard errors are clustered at the census-tract level and reported in parentheses. *, **, and *** denote significance at the 10%, 5%, and 1% levels, respectively.

Table A-8: Wind Direction and Housing Price Effects

	Log Housing Sale Price		
	(1)	(2)	(3)
$\leq$ 5 km × Post	-0.125^{*}	-0.009	-0.064^{***}
	(0.075)	(0.021)	(0.020)
$\leq$ 5 km × Post × DM			-0.002
			(0.003)
Transaction Controls	Yes	Yes	Yes
Property Controls	Yes	Yes	Yes
DC × Distance FE	Yes	Yes	Yes
DC × Census Tract FE	Yes	Yes	Yes
DC × County × Time FE	Yes	Yes	Yes
Adj. R^2	0.660	0.661	0.662
Observations	3,441,999	3,486,749	3,661,868
Sample	Treated: AD Control: All	Treated: ND Control: All	All

Notes: This table presents ordinary least squares estimates from estimating Equation (2). The dependent variable is the natural logarithm of the residential transaction price adjusted to 2010 dollars. The explanatory variable of interest is $\leq$ 5 km × Post, which equals one for transactions of properties located within 5 km of the first-built hyperscale data center occurring after the facility is built. "DM" indicates the number of calendar months the land parcels are in the downwind/upwind sector of the nearby data center. "AD" and "ND" indicate that land parcels are always and never in the downwind sector of the nearby data center throughout a calendar year. The sample is restricted to residential transactions within 50 km of the relevant facility, and observations in the 5–20 km buffer ring are excluded. Columns (1) and (2) further restrict the treatment sample to transactions located on land parcels that are always in the downwind sector and always in the upwind sector, respectively, relative to the nearby data center. Transaction controls include indicators for cash purchases, resale transactions, short sales, foreclosure REO transactions, corporate buyers, local buyers, corporate sellers, local sellers, and whether the seller previously purchased the unit using cash, through a short sale, or through a foreclosure REO transactions. Property controls include an indicator for mobile-home transactions, property type (e.g., single-family, duplex), and indicators for views and key structural attributes (e.g., fuel type, water/utility/sewer type, electrical wiring, roof type and cover, building quality and improvement condition, heating type, floor type, basement, pool), as well as percentiles of number bedrooms, bathrooms, property age, land area, living area, number of stories, and parking. All categorical covariates include an additional missing category. Distance fixed effects control for property–facility distance using 100 bins of width 0.1 km. Standard errors are clustered at the census-tract level and reported in parentheses. *, **, and *** denote significance at the 10%, 5%, and 1% levels, respectively.