

# DIGITAL MONOCENTRIC CITIES: DOMAIN NAME PRICES, SEMANTIC DISTANCE, AND THE ECONOMICS OF ONLINE LOCATION

LUIS ALEJANDRO RUBIANO GUERRERO

Facultad de Economía, Universidad de Los Andes

CARLOS ANDRÉS CASTILLO CABRERA

Facultad de Economía, Universidad de Los Andes

ANDRÉS FELIPE ROSAS CASTILLO

Facultad de Economía, Universidad de Los Andes

We study whether domain name markets exhibit a digital analogue of the monocentric city model, in which “rents” decline with distance from a central location. We conceptualize online location using embedding-based semantic distance from a keyword center (e.g., *chat.com*) and examine how domain prices vary along this virtual center–periphery dimension. Our analysis proceeds in two parts. First, leveraging the release of ChatGPT-3.5 in November 2022 as an exogenous attention shock to the keyword “chat,” we estimate a difference-in-differences event-study with domain and time fixed effects. We find a post-release steepening of the price–distance gradient: semantically central domains experience disproportionately larger price gains. Second, across multiple high-salience keyword clusters, we document consistently negative cross-sectional gradients, with suggestive evidence that more popular categories exhibit steeper slopes. These results imply that search frictions and cognitive costs generate economically meaningful spatial structure in cyberspace, extending core urban equilibrium insights to digital attention markets. **JEL codes:** R12, L86, C23, D83

**KEYWORDS:** Monocentric digital city, Domain names, Semantic distance, Attention shocks.

## 1. INTRODUCTION

The pricing of domain names has become a central element of the digital economy, shaping online visibility, competition, and the distribution of economic activity across virtual space. Premium internet domains have long commanded striking prices, for example *Sex.com* reportedly sold for about \$13 million in 2010 (The Economist, 2010), and *Voice.com* for \$30 million in 2019 (Allemann, 2019). Such eye-popping sales of “virtual real estate” invite an intriguing question: do online domain markets obey economic forces analogous to those in physical land markets? In particular, this paper asks whether internet domain name markets follow a digital version of the classic monocentric city model from urban economics.

Urban economists have long emphasized the importance of location. In the classic monocentric city framework, land rents rise sharply as distance to the central business district (CBD) falls (Alonso, 1964; Mills, 1972; Muth, 1969). Proximity to the city center yields greater access to jobs and amenities, so households and firms are willing to pay a premium for central locations. Modern extensions of this theory continue to find steep spatial gradients in land value (Lucas & Rossi-Hansberg, 2002; Ahlfeldt, Redding, Sturm, & Wolf, 2015). If digital domains exhibit analogous “central” locations in an economic sense, their prices might form a similar gradient with respect to a conceptual center. Examining this possibility is not only novel but

---

Luis Alejandro Rubiano Guerrero: [la.rubiano@uniandes.edu.co](mailto:la.rubiano@uniandes.edu.co)

Carlos Andrés Castillo Cabrera: [ca.castillo@uniandes.edu.co](mailto:ca.castillo@uniandes.edu.co)

Andrés Felipe Rosas Castillo: [a.rosasc@uniandes.edu.co](mailto:a.rosasc@uniandes.edu.co)

also relevant for urban economics: it tests whether spatial equilibrium concepts apply when physical distance is replaced by semantic distance in an online environment.

There is a rich literature debating whether the digital revolution diminishes the importance of geography. Early commentators predicted a “death of distance,” speculating that advances in communication technology would weaken cities and render physical location less relevant (Pascal, 1987; Cairncross, 1997). Subsequent research, however, found that agglomeration forces remain potent: information technology did not eliminate the value of physical proximity or the clustering of economic activity (Moretti, 2021; Greenstone, Hornbeck & Moretti, 2010 ; Gaspar & Glaeser, 1998; Tranos & Nijkamp, 2013). By analogy, distance may not vanish in cyberspace either. Instead, it may reappear in new dimensions. Even in seemingly borderless online markets, user behavior reflects persistent frictions of similar nature as distance: for example, internet consumption tends to favor local or familiar content, indicating that cultural and linguistic “distances” still matter (Blum & Goldfarb, 2006). Moreover, in the digital attention economy, consumer attention is a scarce resource, and being easy to find or remember confers a significant advantage (Ellison & Ellison, 2009). This suggests that a domain name closely aligned with a popular keyword, e.g. `chat.com` in the chatbot context, could attract disproportionate user attention by reducing search and navigation costs. In short, being semantically central (e.g. a one-word generic domain that directly represents a concept) may yield the same kind of rent premium online that being centrally located does in a city.

Recent work provides preliminary evidence that such virtual location effects are real. Lindenthal (2018) adapts the monocentric city model to the domain name market, arguing that an analogous rent gradient exists in cyberspace. He notes that “downtown” domains with highly common or short words are known to transact for millions, whereas domains in the semantic periphery trade for far less. His empirical analysis shows that the supply of these attractive virtual locations is effectively constrained: registration patterns confirm that desirable domains are scarce. Almost all simple, memorable `.com` names have been claimed, and the pool of unregistered high-quality domains has dwindled. This scarcity is reflected in market trends: the rate of new `.com` registrations has slowed even as internet usage grows, and resale prices of existing domains rose by over 60% between 2006 and 2012 (Lindenthal, 2014). Together, these patterns indicate that demand for central domain names outstrips supply, just as limited land in a CBD leads to high rents.

We formalize the idea that user search and attention costs act like transportation costs in a city. In our framework, consumers have limited attention and will more readily navigate to websites that are easier to discover or recall. A domain name closer to a category’s focal keyword, for instance, `chat.com` relative to an obscure name like `conversational-agent.net`, reduces the cognitive effort required for users to find or remember it. Consequently, such a domain will attract more direct traffic or search hits, increasing its revenue potential. Firms and investors are therefore willing to pay a higher “rent” for semantically central domains. This logic yields two key predictions. First, within any given semantic category, domain prices will decline as the domain’s linguistic or thematic distance from the central keyword increases (a negative semantic distance gradient, analogous to a bid-rent curve in urban land markets). Second, a positive shock to overall demand for the central keyword should disproportionately raise the values of domains close to that keyword, steepening the price-distance gradient. In other words, if the “central business district” of a digital market becomes more valuable, we expect the premium on proximity (semantic closeness) to that center to increase.

Our empirical approach proceeds in two parts. First, we exploit the public release of Chat-GPT in November 2022 as a plausibly exogenous shock to the salience of the keyword “chat.” This event provides a setting in which a specific semantic category experiences a sudden, large increase in global attention. As shown in 1, global Google search interest in “chat” remained

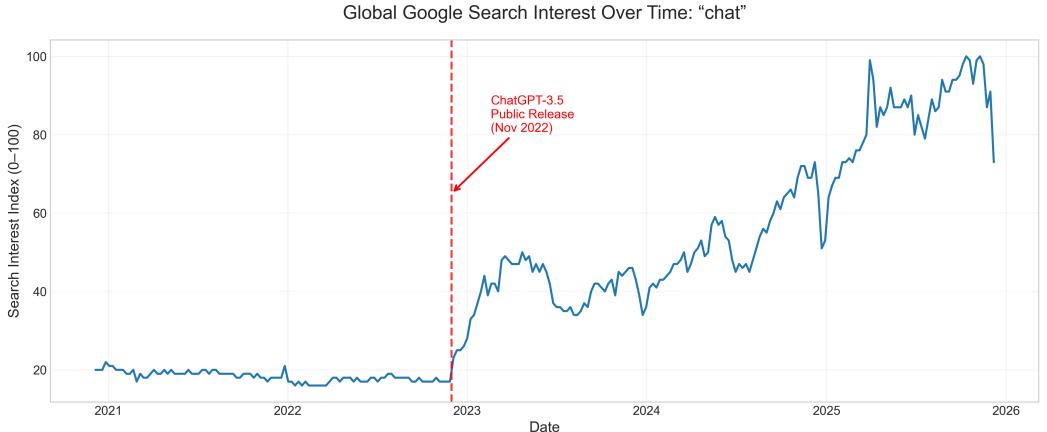


FIGURE 1.  
Global Google Search Interest for the Keyword “chat,” 2021–2026.  
The dashed line marks the public release of ChatGPT-3.5 in November 2022

stable for years before rising sharply and discontinuously at the moment of the ChatGPT-3.5 release, highlighting the sudden increase in attention to this semantic category. We treat the domain `chat.com` as the central location of this semantic “market” and measure each related domain name’s distance from this center using cosine distance in a high-dimensional embedding space. We then estimate how the relationship between domain prices and semantic distance evolves around the ChatGPT release. A difference-in-differences event-study design with domain and time fixed effects allows us to isolate changes in the slope of the price–distance gradient. The evidence indicates a sharp, discontinuous steepening of the gradient immediately after November 2022: domains closest to the semantic center exhibit the largest relative price gains, while those further away show considerably smaller changes. Pre-trend tests reveal no differential evolution in the gradient prior to the release, strengthening the interpretation that the shift arises from the shock.

In the second part of the analysis, we examine whether the structure observed in the “chat” domain cluster generalizes across a broader set of economically salient keywords. We assemble a dataset of exact-match domains for various categories, such as lawyers, finance, hotel, and crypto, along with semantically related domain names for each cluster. For each keyword  $k$ , the canonical `keyword.com` domain serves as the center, and embedding-based semantic distances trace the topology of the surrounding domain space. Across categories, we find robust evidence of declining price gradients: domain values fall steadily with greater semantic distance from the category center, closely mirroring the monotone bid-rent patterns characteristic of monocentric urban models. Moreover, clusters with higher underlying keyword popularity, as proxied by search frequency, exhibit steeper gradients, consistent with models in which larger or more productive “cities” command higher central rents. Taken together, the results suggest that agglomeration forces and spatial price gradients emerge systematically in virtual domain markets.

This paper makes several contributions to the economics of digital markets and to urban economic theory. First, it demonstrates that canonical spatial concepts such as distance, centrality, agglomeration, and bid-rent gradients can emerge endogenously even in environments where physical space plays no role. By showing that domain prices decline predictably with semantic distance from a keyword center, we provide evidence that search frictions and cognitive costs

generate a form of virtual geography. Second, the paper introduces a rigorous, theoretically grounded measure of “location” in digital markets by constructing semantic distances using embedding-based metrics. This approach allows us to map domain names into a continuous, economically meaningful space where proximity reflects similarity in user search behavior. Third, our use of the ChatGPT launch as an exogenous demand shock parallels the identification strategies used in urban and labor economics to study agglomeration effects. Here, it enables us to show that increases in keyword salience concentrate value disproportionately near the semantic center, steepening the digital bid-rent curve. This provides rare causal evidence of virtual agglomeration effects.

The remainder of the paper is organized as follows. Section 1 has introduced the motivating framework and outlined why domain name markets offer a compelling setting in which to test for digital monocentric structures. Section 1 also highlighted the analogy between semantic distance online and physical distance in urban models. Section 2 details the empirical strategy, covering the causal framework, identification design around the ChatGPT shock, and the econometric specifications used to estimate price-distance gradients within and across semantic clusters. Section 3 describes the domain price data, the construction of embedding-based distance measures, and the descriptive patterns motivating the analysis. Section 4 presents the main results: the causal response of the price gradient to the ChatGPT shock and the broader cross-sectional evidence on digital bid-rent curves. Section 5 concludes by discussing implications for urban theory, digital attention markets, and the economic geography of the internet.

## 2. RESEARCH DESIGN

### 2.1. Causal Framework

Our empirical strategy consists of two complementary components. The first component exploits the public release of ChatGPT in November 2022 as a quasi-experimental shock to the economic value of the “chat” semantic category. This event generated a sudden increase in global interest in chatbot-related concepts, which we treat as an exogenous upward shift in demand for domains proximate to the canonical keyword **chat**. For domains  $i$  with semantic distance  $d_i$  from the center (`chat.com`), we formalize potential outcomes as  $Y_{it}(post, d_i)$ , the log price domain  $i$  would take at time  $t$  given its exposure to the pre- or post-shock environment. The causal estimand of interest is the change in the marginal effect of semantic distance on prices after the shock:

$$\tau = \left[ \frac{\partial \mathbb{E}[Y_{it}(1, d_i)]}{\partial d_i} \right] - \left[ \frac{\partial \mathbb{E}[Y_{it}(0, d_i)]}{\partial d_i} \right].$$

A negative value of  $\tau$  indicates that the price-distance gradient steepens after the shock, consistent with an increase in the premium associated with semantic centrality.

The second component of the analysis is observational and focuses on the broader structure of domain markets across multiple high-salience semantic categories. Here, we examine whether domain name prices exhibit stable negative gradients with respect to semantic distance from their respective keyword centers (e.g., `lawyers.com`, `hotel.com`, `crypto.com`). Let  $k$  index a semantic category, and let  $d_{ik}$  denote the semantic distance between domain  $i$  and the central domain for category  $k$ . The observational estimand is the cross-sectional slope

$$\delta_k = \frac{\partial \mathbb{E}[Y_{ik}]}{\partial d_{ik}},$$

which captures the extent to which prices decline as domains become semantically more peripheral within category  $k$ . These two components, one causal, one structural, jointly assess whether domain markets systematically obey a digital analogue of the monocentric city model.

## 2.2. Identification Strategy

### *Part I: Quasi-Experimental Identification Around the ChatGPT Shock*

For the causal component, identification comes from comparing the pre- and post-shock evolution of domain prices as a function of semantic distance from `chat.com`. Domains closer to the center face higher exposure intensity, whereas domains further away face weaker exposure. This yields a difference-in-differences framework with heterogeneous treatment intensity. The key identifying assumption is that, absent the ChatGPT release, the price–distance gradient for “chat-related” domains would have evolved in parallel across different values of  $d_i$ . We assess this assumption through an event-study design that allows us to inspect differential pre-trends in gradient slopes. Additionally, we use domains from unrelated semantic categories as a comparison group to rule out contemporaneous market-wide shocks.

### *Part II: Observational Identification Across Semantic Clusters*

The cross-sectional component is descriptive, but it is grounded in a well-defined economic model. For each semantic category  $k$ , we treat the canonical domain (e.g., `lawyers.com`) as the economic center and measure how prices vary with semantic distance. This relies on the assumption that semantic distance is a valid proxy for user search frictions and that domain prices reflect expected economic returns from direct traffic, memorability, and brand relevance. While the observational regressions do not identify causal effects, the central question is structural: whether the price-distance gradients across categories resemble the bid-rent curves predicted by the monocentric model. To strengthen interpretability, we compare gradients across categories with varying popularity levels, analogous to differences in city size or agglomeration strength.

## 2.3. Estimation Strategy

### *Part I: Treatment Effect of the ChatGPT Shock*

Our baseline specification estimates how the slope of the price–distance curve changes after the ChatGPT launch:

$$\log P_{it} = \alpha_i + \lambda_t + \beta_1 d_i + \beta_2 Post_t + \beta_3 (Post_t \times d_i) + \varepsilon_{it},$$

where  $\alpha_i$  captures time-invariant domain characteristics and  $\lambda_t$  absorbs aggregate market conditions. The coefficient of interest,  $\beta_3$ , measures the differential response of prices to semantic distance in the post-shock period. To assess dynamics, we estimate an event-study model:

$$\log P_{it} = \alpha_i + \lambda_t + \sum_{k \neq -1} \theta_k (d_i \times \mathbf{1}\{t = k\}) + \varepsilon_{it},$$

where the  $\theta_k$  coefficients trace the evolution of the gradient around the shock. Lack of pre-trend in  $\theta_k$  for  $k < 0$  supports the identification strategy.

## Part II: Observational Identification Across Semantic Clusters

For each semantic category  $k$ , we estimate the structural price gradient using:

$$\log P_{ik} = \gamma_k + \delta_k d_{ik} + \mathbf{X}'_{ik} \psi + \xi_{ik},$$

where  $d_{ik}$  captures semantic distance,  $\gamma_k$  is a category fixed effect, and  $\mathbf{X}_{ik}$  includes controls such as domain length, TLD and search volume. A negative  $\delta_k$  indicates a declining bid-rent-like gradient. We then relate these gradients to keyword popularity to test whether more “central” or widely searched concepts exhibit steeper gradients.

### 2.4. Robustness and Specification Checks

We perform several robustness checks for both components of the analysis, these are present in the Appendix.

#### Part I: Causal Component

1. Placebo shocks: We assign pseudo-shock dates in the pre-ChatGPT period and confirm no spurious steepening.
2. Placebo categories: Keywords unrelated to “chat” show no comparable structural break in their gradients.
3. Alternative time windows and bandwidths: Narrow and wide pre/post windows yield consistent patterns in gradient steepening.

#### Part II: Observational Component

1. Cross-category consistency: We verify that gradients remain negative for every traditionally high-value category (e.g., legal, finance, hotel, crypto).
2. Functional form robustness: We estimate log-distance and quadratic distance all yielding similar monotone patterns.
3. Within-category heterogeneity tests: Splits by TLD, domain age, and length confirm stable negative gradients across subgroups.

## 3. DATA

### 3.1. Data-Generating Process

This section describes the data-generating process used to construct the datasets for the two empirical components of the paper: (i) a quasi-experimental panel around the ChatGPT release in the *chat* cluster, and (ii) an observational cross-section spanning multiple keyword clusters.

#### Units of observation.

The fundamental unit is a *domain name*. Domains are organized into semantic clusters indexed by a keyword  $k \in \{\text{chat, lawyers, hotel, finance, crypto, travel}\}$ , where the canonical exact match domain  $k.\text{com}$  is the cluster center. The master list contains one *chat* cluster with 240 non-center variants plus the center, and five non-*chat* clusters with 180 non-center variants plus each center, yielding  $241 + 5 \times 181 = 1,146$  domains prior to any

sale selection. Domain strings are generated to resemble marketplace listings (e.g., adjective–keyword composites, hyphens, digits) and top-level domains (TLDs) are drawn with a realistic tilt toward `.com`.<sup>1</sup>

#### *Time structure.*

The quasi-experimental component is a monthly panel from January 2021 to December 2024 (48 months). The observational component is cross-sectional (one observation per domain), constructed for the same master list of domains.

#### *Semantic distance (“virtual location”).*

A core contribution of our data construction is a continuous measure of semantic distance between domains. We embed each domain string using the Qwen3–Embedding–8B model, one of the strongest publicly available embedding architectures according to recent large-scale benchmark evaluations (Enevoldsen et al., 2025), and compute cosine similarity between domain  $i$  and its category center  $c(k)$ . We then transform similarity into an interpretable distance measure:

$$d_{ik} = \frac{1 - s_{ik}}{0.6},$$

where  $s_{ik}$  is the cosine similarity between domain  $i$  and the canonical center of keyword  $k$ . This normalization maps the empirical range of high-dimensional text-embedding similarities into a scale that is approximately comparable across categories and magnitudes of semantic proximity. The scaling parameter is motivated by evidence that cosine similarities for standard English text embeddings are typically positive and rarely close to zero, implying that raw similarity scores can be compressed in a narrow, right-shifted range (Garg, 2022). Standardizing by 0.6 yields distances that better spread domains along the center–periphery dimension relevant for our bid–rent interpretation.

#### *Treatment assignment mechanism (quasi-experimental component).*

The quasi-experiment exploits a time discontinuity at the public release of ChatGPT (cutoff  $t_0 = \text{November 2022}$ ). Treatment is not a binary assignment at the unit level; instead, domains differ in exposure intensity as a function of distance to the `chat` center. Let  $Post_t = 1\{t \geq t_0\}$ . An attention shifter  $A_t$  follows a “fuzzy” jump in mean at  $t_0$ :

$$A_t \sim \mathcal{N}(\mu_{\text{pre}}, \sigma_A^2) \quad \text{for } t < t_0, \quad A_t \sim \mathcal{N}(\mu_{\text{post}}, \sigma_A^2) \quad \text{for } t \geq t_0,$$

clipped to a bounded support. The baseline calibration sets  $\mu_{\text{pre}} = 0.35$ ,  $\mu_{\text{post}} = 0.75$ , and  $\sigma_A = 0.08$ . Domain-level exposure is then

$$T_{it} = A_t \exp(-\kappa d_i),$$

with  $\kappa = 2.2$ , so that proximity to `chat.com` implies higher exposure to the post-2022 demand shock.

---

<sup>1</sup>For the purposes of this course project, the underlying dataset is simulated to follow the Urban Economics Final project guidelines. The data is made to simulate datasets that platforms such as GoDaddy and Semrush Traffic Analytics may have, we compute distances using Qwen3–Embedding–8B.

### *Outcome variable and functional form.*

The main outcome is log price. We distinguish (i) a latent log value (economic fundamentals) and (ii) an observed transaction price that is only recorded upon sale.

(i) *Quasi-experimental chat panel.* For domain  $i$  and month  $t$ , latent log price is generated as

$$\log P_{it}^* = \underbrace{c_0}_{\text{baseline}} + \underbrace{\alpha_i}_{\text{domain FE}} + \underbrace{\lambda_t}_{\text{time FE}} + \underbrace{s_t d_i}_{\text{distance gradient}} + \underbrace{\rho T_{it}}_{\text{attention channel}} + \mathbf{X}'_i \beta + \varepsilon_{it},$$

where  $s_t = -(g_0 + g_1 \cdot Post_t)$  allows the price–distance slope to steepen after  $t_0$ . The calibration uses  $g_0 = 1.1$  and  $g_1 = 0.7$ , implying a more negative gradient post-shock;  $\rho = 0.45$ , and  $\varepsilon_{it} \sim \mathcal{N}(0, \sigma_{\text{chat}}^2)$  with  $\sigma_{\text{chat}} = 0.22$ .

(ii) *Cross-sectional multi-keyword sample.* For each domain  $i$  in cluster  $k$ , latent log price is

$$\log P_{ik}^* = \gamma_k + \alpha_i + \delta_k d_{ik} + \mathbf{X}'_{ik} \psi + u_{ik},$$

where  $\gamma_k$  is a cluster intercept and  $\delta_k < 0$  is a cluster-specific gradient. Cluster “size” (popularity)  $S_k$  is generated as a bounded normal draw and maps into the gradient via

$$\delta_k = -(b_0 + b_1 S_k),$$

so more popular keywords have steeper bid-rent-like curves. The calibration uses  $b_0 = 1.0$  and  $b_1 = 0.9$ .

### *Covariates.*

The vector  $\mathbf{X}$  contains domain characteristics commonly associated with valuations: an indicator for `.com`, length of the second-level domain, age (in years), and indicators for hyphens and digits (plus an exact-match indicator used for normalization). Age is generated as a clipped lognormal draw to reflect skewness in domain ages, and these characteristics enter log prices through a standard hedonic form (log transforms for length and age, level shifts for `.com`/hyphen/digit). These covariates are intentionally correlated with semantic distance in the domain-construction step (e.g., longer names and non-`.com` TLDs are more prevalent away from the center), mirroring the empirical fact that “peripheral” domains tend to be longer, less memorable, and less likely to be `.com`.

### *Fixed effects.*

The *chat* panel includes domain fixed effects  $\alpha_i \sim \mathcal{N}(0, 0.35^2)$  and month fixed effects  $\lambda_t \sim \mathcal{N}(0, 0.05^2)$  to represent time-invariant domain quality and market-wide pricing cycles, respectively. In the cross-section,  $\gamma_k$  serves as a cluster intercept, while  $\alpha_i$  captures idiosyncratic quality.

### *Error structure.*

Idiosyncratic errors are Gaussian in the panel, and in the cross-section we allow rare, large “auction-like” shocks through a low-probability additive component (a heavy-tailed mixture), reflecting occasional extreme sales in domain markets. The baseline probability of a tail shock is 1.5%.

*Complications: selection into observed sales (missing outcomes).*

A key complication is that transaction prices are only observed when a domain is sold through a marketplace/registry. We model this as endogenous selection: conditional on fundamentals, higher-value domains are more likely to transact. Specifically, we generate a sale indicator  $Sold \in \{0, 1\}$  from a logistic index in latent log price:

$$\Pr(Sold = 1 | \log P^*) = \Lambda(a_0 + a_1 \log P^*),$$

with  $(a_0, a_1) = (-2.2, 0.75)$  in the *chat* panel and  $(-2.0, 0.7)$  in the cross-section. Observed log price is then  $\log P = \log P^* + \eta$  if sold, with  $\eta \sim \mathcal{N}(0, 0.08^2)$  capturing transaction-level noise (negotiation, auction format, reporting).

*Justification of design choices.*

The data generating process is designed to match three empirical regularities of domain markets relevant for our research question. First, it embeds a monotone price gradient in semantic distance, and allows that gradient to steepen after a large demand shock (the ChatGPT release) through the interaction of  $Post_t$  with distance. Second, it incorporates `.com` and penalties for length/hyphens/digits consistent with the idea that memorability and brandability raise value, which is central to interpreting semantic proximity as a reduction in “navigation costs.” Third, it models the fact that researchers typically observe prices only for transacted domains, and that the propensity to transact is increasing in value.

### 3.2. Descriptive Statistics

Descriptively, the data exhibit large dispersion in transaction values within each semantic cluster, with the highest prices concentrated in short, exact-match, and older domains. Across categories, canonical *keyword.com* centers occupy the upper tail of the price distribution, while semantically peripheral domains trade at substantially lower prices. In the *chat* panel, we observe a pronounced upward shift in transaction values for domains close to `chat.com` after the public release of ChatGPT, consistent with a demand-driven revaluation of the semantic center and its immediate neighborhood. In the multi-category cross-section, clusters associated with higher baseline keyword popularity display systematically steeper negative relationships between price and embedding distance.

Table I documents the basic properties of the *chat* panel. The panel contains 2,372 domain-period observations, and the average value of the post-ChatGPT indicator  $post_{it}$  is approximately 0.52, implying that the sample is roughly balanced between pre- and post-treatment observations. The exogenous attention measure  $attention\_index_t$  has a mean of about 0.53 and ranges from 0.15 to 0.88, indicating substantial time-series variation in aggregate interest in “chat”-related domains. The exposure index  $T_{it}$ , which combines attention and distance, has a mean of 0.26 and an interquartile range from 0.16 to 0.36. This confirms that the quasi-experimental shock interacts meaningfully with cross-sectional heterogeneity in virtual location.

The semantic distance measures reveal a wide spread in how close domains are to the center of the *chat* cluster. Both the embedding-based distance and the “true” distance have means around 0.33 and 0.31, respectively, with interquartile ranges roughly between 0.25 and 0.42. A non-trivial share of domains lies quite close to the semantic center, while the upper tail extends to distances around 0.8.

The remaining covariates describe a sample of relatively standard but commercially oriented names. About three quarters of the observations correspond to `.com` domains, and the median

TABLE I  
DESCRIPTIVE STATISTICS: CHAT PANEL SAMPLE

Variable	<i>N</i>	Mean	Min	25th pctl	Median	75th pctl	Max
Post-ChatGPT indicator ( $post_{it}$ )	2,372	0.524	0.000	0.000	1.000	1.000	1.000
Attention index ( $attention\_index_t$ )	2,372	0.533	0.146	0.335	0.598	0.741	0.881
Embedding distance ( $embedding\_distance$ )	2,372	0.336	-0.000	0.250	0.330	0.415	0.786
True distance ( $true\_distance$ )	2,372	0.306	0.000	0.171	0.303	0.436	0.785
Distance used ( $distance\_used$ )	2,372	0.336	0.000	0.250	0.330	0.415	0.786
Exposure index ( $T_{it}$ )	2,372	0.264	0.026	0.155	0.245	0.358	0.861
.com indicator	2,372	0.744	0.000	0.000	1.000	1.000	1.000
Length (characters)	2,372	11.656	4.000	9.000	11.000	14.000	20.000
Age (years)	2,372	3.933	0.689	2.179	3.438	4.648	30.000
Hyphen indicator	2,372	0.145	0.000	0.000	0.000	0.000	1.000
Digit indicator	2,372	0.153	0.000	0.000	0.000	0.000	1.000
Exact-match indicator	2,372	0.006	0.000	0.000	0.000	0.000	1.000
Domain effect ( $\alpha_i$ )	2,372	0.076	-0.967	-0.104	0.118	0.289	0.941
Time effect ( $\lambda_t$ )	2,372	0.011	-0.091	-0.029	0.022	0.046	0.089
Latent log price ( $\log p^{\text{latent}}$ )	2,372	1.194	-0.436	0.852	1.221	1.535	2.499
Sale probability ( $sale\_prob$ )	2,372	0.220	0.074	0.173	0.217	0.260	0.419
Sale indicator ( $sold$ )	2,372	1.000	1.000	1.000	1.000	1.000	1.000
Observed log price ( $\log p^{\text{obs}}$ )	2,372	1.195	-0.411	0.837	1.226	1.545	2.574
Observed price ( $p^{\text{obs}}$ )	2,372	3.712	0.663	2.309	3.406	4.687	13.120

*Notes:* This table reports descriptive statistics for the chat panel, which comprises 2,372 domain-period observations for domains in the “chat” semantic cluster. The variables  $post_{it}$  and  $attention\_index_t$  capture the post-ChatGPT period and the exogenous evolution of aggregate attention, respectively. Distance measures summarize semantic distance from each domain to the cluster center. The variables  $T_{it}$ ,  $\alpha_i$ , and  $\lambda_t$  correspond to the exposure index and the domain and time effects that shape latent log prices and sale probabilities. Observed prices and log prices correspond to realized sales in the panel; by construction, all panel observations are sold ( $sold = 1$ ).

TABLE II  
DESCRIPTIVE STATISTICS: MULTI-KEYWORD CROSS-SECTION

Variable	<i>N</i>	Mean	Min	25th pctl	Median	75th pctl	Max
.com indicator	1,146	0.717	0.000	0.000	1.000	1.000	1.000
Length (characters)	1,146	13.012	4.000	11.000	13.000	15.000	24.000
Has hyphen	1,146	0.147	0.000	0.000	0.000	0.000	1.000
Has digit	1,146	0.201	0.000	0.000	0.000	0.000	1.000
Token count	1,146	1.194	1.000	1.000	1.000	1.000	3.000
Exact-match indicator	1,146	0.005	0.000	0.000	0.000	0.000	1.000
Age (years)	1,146	4.054	0.475	2.314	3.422	4.956	30.000
True distance ( $true\_distance$ )	1,146	0.328	0.000	0.191	0.317	0.453	0.923
Embedding distance ( $embedding\_distance$ )	1,146	0.336	-0.000	0.237	0.332	0.426	0.786
Distance used ( $distance\_used$ )	1,146	0.336	0.000	0.237	0.332	0.426	0.786
Cluster attention ( $S_k$ )	1,146	0.615	0.229	0.549	0.710	0.736	0.750
Cluster slope ( $slope_k$ )	1,146	-1.553	-1.675	-1.662	-1.639	-1.494	-1.206
Domain effect ( $\alpha_i$ )	1,146	-0.007	-1.277	-0.234	0.003	0.214	1.113
Cluster shifter ( $\gamma_k$ )	1,146	1.686	1.138	1.614	1.723	1.799	2.017
Latent log price ( $\log p^{\text{latent}}$ )	1,146	1.265	-0.383	0.880	1.235	1.626	3.769
Sale probability ( $sale\_prob$ )	1,146	0.254	0.094	0.200	0.243	0.297	0.654
Sale indicator ( $sold$ )	1,146	0.257	0.000	0.000	0.000	1.000	1.000
Observed log price ( $\log p^{\text{obs}}$ )	294	1.441	0.267	1.016	1.416	1.775	3.660
Observed price ( $p^{\text{obs}}$ )	294	5.129	1.306	2.763	4.122	5.903	38.844

*Notes:* This table reports descriptive statistics for the multi-keyword cross-section, which comprises 1,146 domains across several economically salient keyword clusters. The variables  $S_k$ ,  $slope_k$ , and  $\gamma_k$  summarize cluster-level attention, the steepness of the price-distance gradient, and a cluster-specific shifter, respectively. Latent log prices and sale probabilities are defined at the domain level, while observed prices and log prices are only available for the 294 domains that sell at least once over the sample window ( $sold = 1$ ).

length is 11 characters, with an interquartile range from 9 to 14 characters. Hyphens and digits are relatively rare (means of 0.15 and 0.15, respectively), and exact matches to the focal keyword are almost non-existent (around 0.6 percent of observations). Domains are relatively young, with a median age of roughly 3.4 years and an upper quartile of about 4.6 years, although a small subset of domains reaches ages of up to 30 years, consistent with the presence of a long-lived tail of more established assets.

Latent log prices  $\log p^{\text{latent}}$  average around 1.19, with an interquartile range from 0.85 to 1.54. The corresponding sale probability  $sale\_prob$  has a mean of approximately 0.22 and is fairly concentrated between 0.17 and 0.26. By construction, all panel observations correspond to realized transactions, so  $sold = 1$  throughout. Observed log prices  $\log p^{\text{obs}}$  closely track their

latent counterparts, with similar location and dispersion, and imply observed prices  $p^{\text{obs}}$  that are clustered around 3.4 in the median and extend to a maximum of about 13.1. Overall, the chat panel captures a set of moderately priced, mostly .com domains with meaningful variation in attention and virtual location, providing a suitable environment to study the dynamic impact of the ChatGPT shock on domain prices.

Table II turns to the multi-keyword cross-section. The cross-sectional sample consists of 1,146 domains drawn from several high-salience keyword clusters. The composition of basic domain characteristics is similar to the chat panel but slightly more skewed toward longer, potentially more brandable names: the median length is 13 characters, with an interquartile range from 11 to 15, and 72 percent of domains use the .com extension. Hyphens and digits are again relatively uncommon, and exact matches to the focal keyword remain rare (around 0.5 percent of the sample). The age distribution is comparable to the panel, with a median age of 3.4 years and an upper quartile of roughly 5 years, plus a long right tail reaching 30 years.

The distribution of semantic distances in the cross-section mirrors that of the chat panel. Average true and embedding-based distances are around 0.33, with interquartile ranges spanning approximately 0.19 to 0.45. This indicates that the cross-section deliberately covers both domains that are very close to each *keyword.com* center and others that are more peripheral in semantic space. Cluster-level parameters summarize systematic heterogeneity across keyword markets: the attention index  $S_k$  is tightly concentrated between about 0.55 and 0.74, while the slope parameter  $slope_k$  is consistently negative (with a mean around  $-1.55$ ), consistent with the presence of declining prices as semantic distance from the center increases. The shifter  $\gamma_k$  lies between roughly 1.14 and 2.02, capturing level differences in the price schedule across clusters.

Latent log prices in the cross-section are slightly higher and more dispersed than in the chat panel, with a mean of 1.27 and an upper quartile of about 1.63, reflecting the presence of especially valuable domains in some clusters. The implied sale probabilities average around 0.25, somewhat larger than in the panel, and range up to 0.65, suggesting that the cross-section contains a non-trivial share of domains with very high underlying liquidity. Consistent with this, only about 26 percent of domains actually sell ( $sold = 1$ ), but conditional on sale, observed log prices have a mean of 1.44 and reach values above 3.6, corresponding to observed prices that can exceed 38 in the upper tail. Taken together, the cross-sectional statistics highlight a richer upper tail of prices and a greater dispersion in cluster-specific price gradients, complementing the more focused chat panel and enabling a broader assessment of how virtual location shapes domain values across economically salient niches.

Figure 2 summarizes the cross-sectional heterogeneity in prices across keyword clusters. Lawyers stand out as the most expensive category, with mean observed prices clearly above those in finance, chat and hotel, while travel- and crypto-related names occupy the lower end of the distribution. This pattern is consistent with the idea that willingness to pay for domain names reflects underlying profitability and client value in the associated activities, and it motivates allowing for rich cluster-level heterogeneity in the empirical specification.

Figure 3 displays the distribution of domain age in the chat panel. The mass of the distribution is concentrated among relatively young domains between roughly one and five years old, with a long but thin right tail extending toward older, legacy assets. This right-skewed pattern suggests that the panel combines a large stock of recently registered names with a smaller set of long-lived domains that have survived for more than a decade, in line with the view that successful names tend to persist.

Figure 4 shows that the sample is heavily dominated by .com domains, which account for roughly 72% of observations. The remaining share is split across a small set of alternative extensions, including .net, .io, .org, .co and .ai, each of which represents only a single-digit percentage of the sample. This composition reflects the continued centrality of .com as

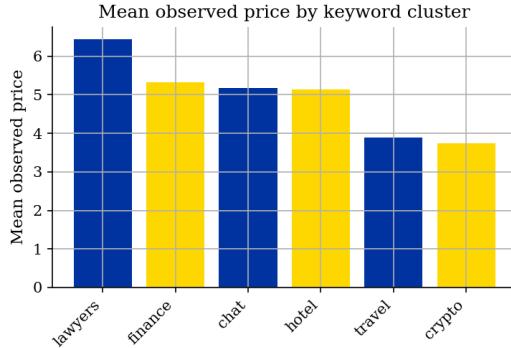


FIGURE 2.  
Mean observed price by keyword cluster

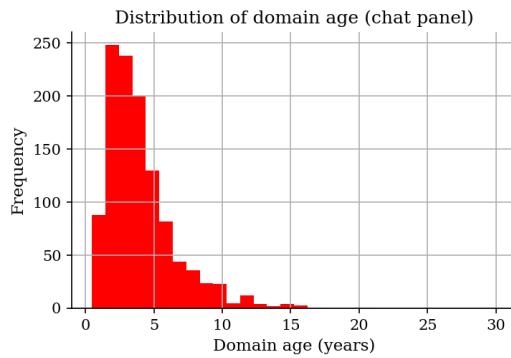


FIGURE 3.  
Distribution of domain age in the chat panel

the benchmark extension in aftermarket trading, with other TLDs playing a secondary but non-negligible role.

Figure 5 plot the distribution of embedding-based semantic distances from each domain to its corresponding keyword center. The distribution is unimodal and concentrated around values between roughly 0.25 and 0.45, with support extending from near zero up to about 0.8. This pattern indicates that the sample spans a wide range of virtual locations: many domains lie relatively close to their canonical *keyword.com*, while a non-trivial fraction occupies more peripheral positions in semantic space. The resulting dispersion in distance is what underpins the bid-rent-style price gradient estimated in the empirical analysis.

Figure 6 plots the evolution of mean observed log prices in the chat panel separately for domains that are relatively close to the semantic center of the cluster (“Near center”) and domains that are farther away (“Far from center”), restricting attention to the pre-ChatGPT period ( $post_{it} = 0$ ). Both series display moderate month-to-month variation but move along broadly similar trajectories, without any systematic divergence or differential pre-trend. In particular, there is no clear evidence of a persistent upward or downward drift in prices for one group relative to the other prior to the treatment date.

This graphical evidence is consistent with the parallel-trends assumption underlying our difference-in-differences interpretation: absent the ChatGPT-induced shift in attention, high-

TLD distribution

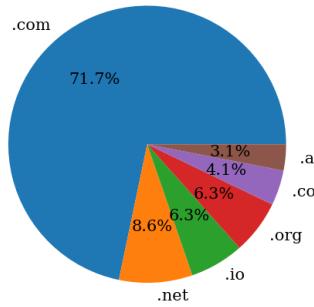


FIGURE 4.  
TLD distribution in the analysis sample

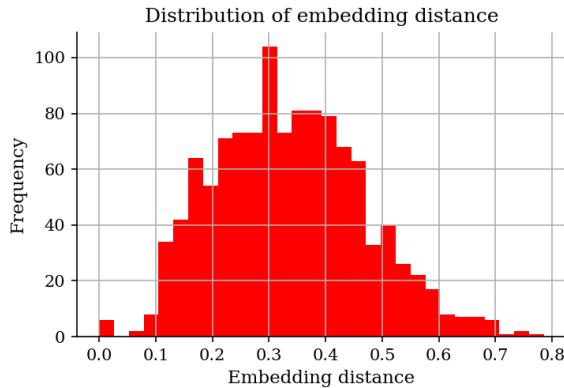


FIGURE 5.  
Distribution of embedding-based semantic distance

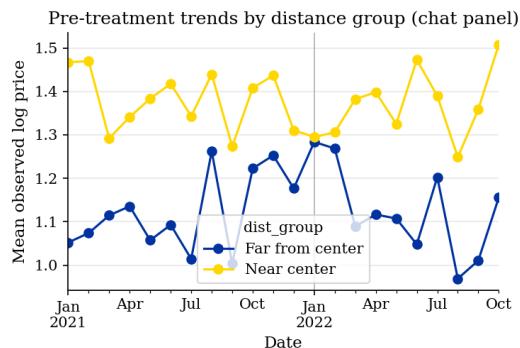


FIGURE 6.  
Pre-treatment trends in mean log prices by distance group (chat panel)

and low-exposure domains would have continued to follow comparable price dynamics. The post-treatment estimates in Section 4 can therefore be interpreted as capturing the interaction between the exogenous change in aggregate attention and virtual location in the semantic space, rather than pre-existing divergent trends between near- and far-from-center domains.

## 4. RESULTS

This section presents the main empirical evidence on whether domain prices exhibit a digital analogue of monocentric bid-rent gradients. We structure the results around the two complementary components of the empirical strategy: (i) the quasi-experimental event-study leveraging the public release of ChatGPT in November 2022 as a shock to the salience of the keyword *chat*, and (ii) the cross-sectional analysis of price-distance gradients across multiple economically salient keyword clusters. The key prediction is that prices decline with semantic distance from a category center and that an exogenous increase in attention to the center should steepen this gradient.

### 4.1. Part I: The ChatGPT shock and the “chat” cluster

Our difference-in-differences framework with heterogeneous treatment intensity estimates whether the marginal effect of semantic distance on log prices shifted after the release date. The coefficient of interest is the interaction between the post indicator and distance, which captures the change in the slope of the price-distance curve.

The baseline estimates indicate a statistically and economically meaningful steepening of the price-distance gradient. In particular, the post  $\times$  distance interaction is negative, implying that domains closer to `chat.com` experienced larger relative price increases than more semantically peripheral names.

To visualize the dynamics of this slope change, we estimate an event-study specification that interacts monthly indicators with semantic distance, using month  $-1$  as the reference period. The resulting coefficients can be interpreted as month-specific shifts in the slope of log price with respect to distance.

Figure 7 shows no evidence of systematic differential pre-trends in the gradient prior to November 2022: the pre-period slope shifts are centered near zero and are statistically indistinguishable from the reference month. Immediately after the ChatGPT release, however, the estimated slope shifts become markedly more negative, indicating a discrete steepening of the digital bid-rent curve.

### 4.2. Part II: Price-distance gradients across keyword clusters

We next examine whether the bid-rent-like structure observed in the *chat* cluster generalizes across other high-salience semantic categories. For each keyword  $k$ , we treat the canonical exact-match domain  $k.com$  as the center and estimate the cross-sectional relationship between observed log prices and embedding-based semantic distance. A negative gradient  $\delta_k$  indicates that domains become less valuable as they move away from the semantic center.

Across categories, the estimated gradients are uniformly negative, indicating that semantic peripherality is associated with lower domain valuations. The magnitude of these slopes varies across clusters, consistent with systematic differences in underlying market size and profitability.

We then test whether keyword popularity is associated with steeper digital gradients, analogous to the prediction that larger or more productive cities exhibit higher central rents and sharper bid-rent curves.

### Event-Study: Change in the Price–Distance Gradient Around ChatGPT

Coefficients are slopes on distance relative to month -1

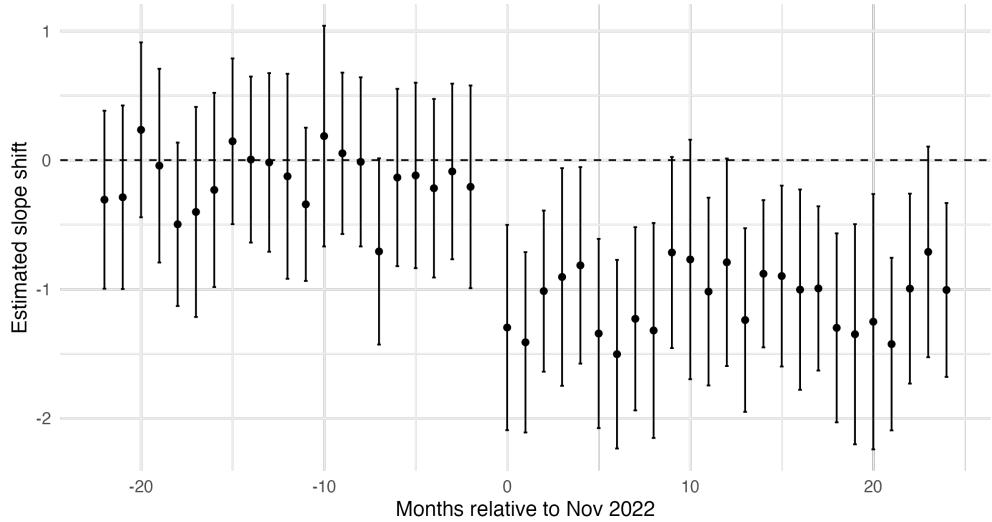


FIGURE 7.

Event-Study: Change in the Price–Distance Gradient Around ChatGPT. Coefficients are slopes on semantic distance relative to month -1.

### Keyword Popularity and Price–Distance Gradient

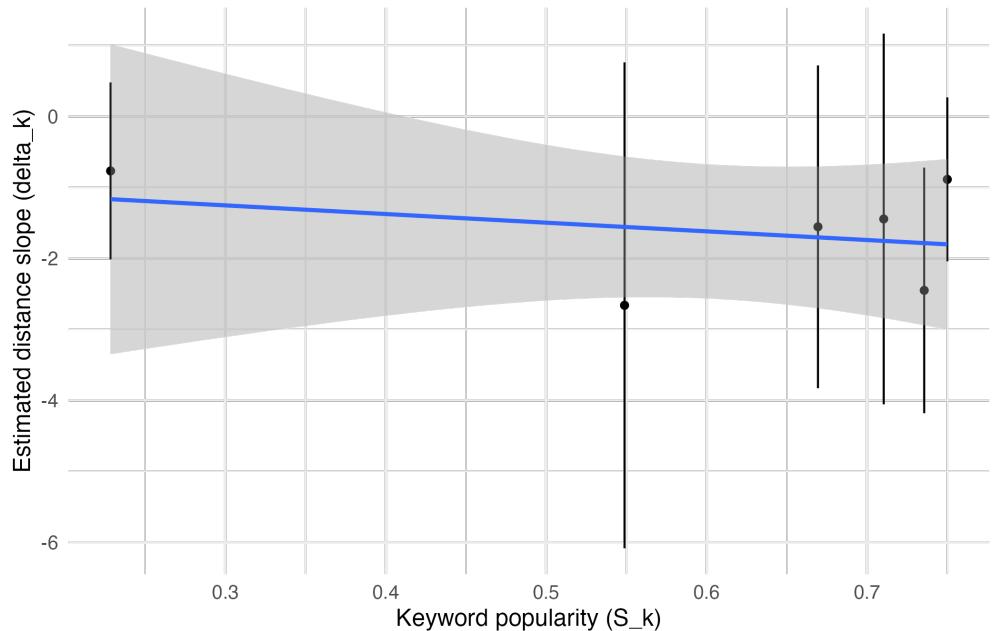


FIGURE 8.—Keyword Popularity and the Price–Distance Gradient. Each point is a keyword-level gradient estimate with confidence intervals.

TABLE III  
CHAT CLUSTER: CHANGE IN THE PRICE–DISTANCE GRADIENT AFTER CHATGPT

	Naive OLS	Domain+Time FE
(Intercept)	1.451*** (0.109)	
distance_used	−1.276*** (0.174)	
post	0.139*** (0.040)	
length	−0.000 (0.007)	
age_years	0.005 (0.007)	
is_com	0.362*** (0.046)	
has_hyphen	−0.187*** (0.059)	
has_digit	−0.103+ (0.053)	
exact_match	0.070 (0.082)	
distance_used × post	−0.817*** (0.114)	
post × distance_used		−0.953*** (0.080)
Num.Obs.	2372	2371
R2	0.336	0.790
R2 Adj.	0.333	0.761
R2 Within		0.058
R2 Within Adj.		0.057
AIC	2401.3	220.2
BIC	2459.0	1882.3
RMSE	0.40	0.22
Std. Errors	by: domain	by: domain
FE: domain		X
FE: date		X

+ $p < 0.1$ , \* $p < 0.05$ , \*\* $p < 0.01$ , \*\*\* $p < 0.001$

Figure 8 displays a negative association between the keyword popularity index  $S_k$  and the estimated distance slope  $\delta_k$ , suggesting that more salient categories command a stronger premium for semantic centrality. Although the cross-sectional design is descriptive, this pattern aligns with the monocentric analogy.

## 5. CONCLUSION

We examine whether domain name markets obey a digital analogue of the monocentric city model, where “rents” fall with distance from a central location, and whether a positive demand shock steepens this gradient. Two findings emerge. First, within the chat cluster, the public release of ChatGPT-3.5 in November 2022 generates a discrete change in the price–distance relationship: the interaction between the post indicator and semantic distance is strongly negative in specifications with domain and time fixed effects, indicating that domains closer to `chat.com` experienced larger relative price gains than more semantically peripheral names.

TABLE IV  
CROSS-SECTIONAL PRICE–DISTANCE GRADIENTS ACROSS KEYWORDS

	Pooled + Category FE	Cluster: chat	Cluster: crypto	Cluster: finance	Cluster: hotel	Cluster: lawyers	Cluster: travel
distance_used	−1.360*** (0.336)	−0.888 (0.590)	−2.451* (0.883)	−2.662 (1.746)	−1.556 (1.160)	−1.446 (1.332)	−0.769 (0.637)
length	0.008 (0.012)	0.036 (0.022)	−0.056* (0.024)	−0.059 (0.052)	0.009 (0.032)	0.006 (0.028)	0.040 (0.048)
age_years	−0.004 (0.010)	−0.007 (0.025)	0.016 (0.027)	0.024 (0.016)	−0.025 (0.029)	−0.024 (0.024)	0.113* (0.053)
is_com	0.417*** (0.071)	0.610*** (0.134)	0.607*** (0.146)	0.193 (0.242)	0.432* (0.180)	0.494** (0.150)	0.155 (0.228)
has_hyphen	0.063 (0.237)	0.089 (0.496)	0.285 (0.473)	−0.091 (1.020)	0.439 (0.355)	−0.168 (0.372)	1.191 (0.725)
has_digit	−0.076 (0.079)	−0.107 (0.168)	−0.140 (0.198)	0.098 (0.346)	−0.081 (0.175)	−0.148 (0.165)	−0.328 (0.216)
token_count	−0.125 (0.154)	−0.170 (0.359)	−0.091 (0.303)	0.021 (0.573)	−0.393+ (0.223)	−0.001 (0.255)	−1.232* (0.595)
exact_match	−0.036 (0.435)	0.837** (0.294)		−1.623+ (0.914)			
(Intercept)		1.154** (0.411)	2.532*** (0.457)	3.026** (0.934)	2.030*** (0.308)	1.830*** (0.429)	1.814** (0.500)
Num.Obs.	294	77	32	45	45	59	36
R2	0.254	0.361	0.507	0.147	0.281	0.184	0.313
R2 Adj.	0.220	0.286	0.363	−0.043	0.145	0.072	0.141
R2 Within	0.189						
R2 Within Adj.	0.165						
AIC	462.6	116.4	39.8	107.1	67.4	97.7	54.5
BIC	514.2	137.5	51.6	123.4	81.8	114.3	67.2
RMSE	0.51	0.46	0.35	0.65	0.43	0.48	0.41
Std. Errors	Heteroskedasticity-robust						
FE: cluster_keyword	X						

+p < 0.1, \*p < 0.05, \*\*p < 0.01, \*\*\*p < 0.001

TABLE V  
RELATIONSHIP BETWEEN KEYWORD POPULARITY AND GRADIENT STEEPNESS

Delta_k on popularity	
(Intercept)	−0.888 (0.875)
S_k	−1.220 (1.414)
Num.Obs.	6
R2	0.096
R2 Adj.	−0.130
AIC	16.4
BIC	16.0
RMSE	0.68
Std.Errors	Heteroskedasticity-robust

+p < 0.1, \*p < 0.05, \*\*p < 0.01, \*\*\*p < 0.001

Second, this pattern is not unique to chat. Across multiple high-salience keyword clusters, we recover consistently negative cross-sectional price–distance gradients, with suggestive evidence that more popular keywords exhibit steeper gradients. Together, these results support the view that online domain markets exhibit a structured center–periphery logic, similar to urban bid-rent curves.

These findings matter for urban theory because they suggest that the monocentric framework is not merely a story about physical commuting costs, it is also a general theory of how willingness to pay organizes around a focal “center” when access, visibility, or search costs are scarce resources. In our setting, semantic distance functions as a substitute for physical distance. The

fact that prices decline smoothly with embedding-based distance across categories, and that the gradient steepens after an exogenous attention shock, implies that core urban mechanisms like concentration of value, and steep rent gradients can emerge when “location” is defined conceptually rather than geographically.

The results also speak directly to digital attention markets. The ChatGPT shock illustrates how sudden increases in aggregate attention can reallocate value toward semantically central assets. Domains that most directly capture the category keyword appear to benefit disproportionately from heightened salience, consistent with the idea that cognitive accessibility and search convenience generate rents in the same way that reduced travel costs do in physical cities. This suggests that digital asset pricing is tightly linked to fluctuating attention fundamentals, and that major technological or cultural shocks can reshape the slope of digital “land value” schedules, not just average prices.

More broadly, the paper contributes to an emerging economic geography of the internet. Even in a setting with negligible physical transport costs, markets do not appear spatially flat. Instead, value organizes along a semantic topology structured by language, memorability, and category salience. In this sense, the internet exhibits a form of geography where the “distance” that matters is the distance in meaning-space. Our evidence indicates that this topology is economically consequential and systematically priced.

Several design insights follow from the two-part strategy. The quasi-experimental component highlights the value of combining time discontinuities with continuous treatment intensity: the event-study of slope shifts provides a direct diagnostic of pre-trends and clarifies that the post-2022 break is driven by a change in the gradient rather than a uniform price jump. The cross-sectional component, while descriptive, strengthens external relevance by showing that the center–periphery structure is not a single-category artifact. Robustness checks using placebo dates, alternative event windows, functional forms, and subgroup splits reinforce the interpretation that the main patterns reflect centrality rents.

Moving from design to implementation would require assembling real transaction and listing data at scale. A feasible empirical path would combine aftermarket sales records, registry-level information on domain age and TLD, and independent measures of keyword salience from search or traffic analytics. With these inputs, the semantic-distance approach could be applied to real domain strings using state-of-the-art embedding models, enabling tests of whether attention shocks systematically steepen digital rent gradients across categories.

The framework nonetheless has limitations. The observational gradients cannot rule out all confounding from correlated domain attributes, and real-world implementation would face measurement challenges in transaction selection, unobserved negotiation dynamics, and heterogeneous market microstructure across TLDs and platforms. Future work could incorporate richer measures of traffic and revenue, explore category-specific competition dynamics, and test how platform algorithms and search ranking systems interact with semantic centrality to amplify digital agglomeration.

## APPENDIX A: ADDITIONAL ROBUSTNESS CHECKS: CHATGPT EVENT-STUDY

This appendix reports supplementary analyses that assess the robustness of the estimated post-ChatGPT steepening in the *chat* price–distance gradient.

## APPENDIX B: CROSS-SECTIONAL SPECIFICATION CHECKS

We complement the baseline cross-sectional gradients with alternative functional forms and subgroup analyses.

TABLE VI  
PART I ROBUSTNESS: PLACEBO SHOCK DATES

	Placebo 2021-06	Placebo 2021-11	Placebo 2022-03
post_pl × distance_used	-0.521*** (0.128)	-0.634*** (0.091)	-0.723*** (0.086)
Num.Obs.	2371	2371	2371
R2	0.779	0.781	0.783
R2 Adj.	0.748	0.751	0.754
R2 Within	0.007	0.018	0.028
R2 Within Adj.	0.007	0.018	0.027
AIC	343.8	317.6	293.9
BIC	2005.9	1979.7	1955.9
RMSE	0.23	0.23	0.23
Std.Errors	by: domain	by: domain	by: domain
FE: domain	X	X	X
FE: date	X	X	X

+ $p < 0.1$ , \* $p < 0.05$ , \*\* $p < 0.01$ , \*\*\* $p < 0.001$

TABLE VII  
PART I ROBUSTNESS: ALTERNATIVE EVENT WINDOWS

	±6m	±12m	±24m
post × distance_used	-0.998*** (0.166)	-0.995*** (0.116)	-0.962*** (0.081)
Num.Obs.	582	1202	2328
R2	0.838	0.800	0.791
R2 Adj.	0.756	0.746	0.761
R2 Within	0.073	0.063	0.059
R2 Within Adj.	0.070	0.062	0.058
AIC	152.8	240.3	224.7
BIC	1008.7	1548.9	1875.7
RMSE	0.20	0.22	0.22
Std.Errors	by: domain	by: domain	by: domain
FE: domain	X	X	X
FE: date	X	X	X

+ $p < 0.1$ , \* $p < 0.05$ , \*\* $p < 0.01$ , \*\*\* $p < 0.001$

## REFERENCES

- ZHANG, YANZHAO, LI, MINGXIN, LONG, DINGKUN, ZHANG, XIN, LIN, HUAN, YANG, BAOSONG, XIE, PENGJUN, YANG, AN, LIU, DAYIHENG, LIN, JUNYANG, HUANG, FEI, AND ZHOU, JINGREN (2025): “Qwen3 Embedding: Advancing Text Embedding and Reranking Through Foundation Models,” *arXiv preprint arXiv:2506.05176*.
- Allemann, A. (2019, June 18). Record breaker: *Voice.com* domain name sells for \$30 million. *Domain Name Wire*. Retrieved from <https://domainnamewire.com/2019/06/18/record-breaker-voice-com-domain-name-sells-for-staggering-30-million/>
- Alonso, W. (1964). *Location and land use: Toward a general theory of land rent*. Cambridge, MA: Harvard University Press.
- Ahlfeldt, G. M., Redding, S. J., Sturm, D. M., & Wolf, N. (2015). The economics of density: Evidence from the Berlin Wall. *Econometrica*, 83(6), 2127–2189.
- Blum, B. S., & Goldfarb, A. (2006). Does the internet defy the law of gravity? *Journal of International Economics*, 70(2), 384–405.

TABLE VIII  
PART II ROBUSTNESS: FUNCTIONAL FORM

	Linear	Quadratic	Log(1 + d)
distance_used	-1.360*** (0.336)	-0.282 (0.953)	
length	0.008 (0.012)	0.005 (0.013)	0.008 (0.013)
age_years	-0.004 (0.010)	-0.003 (0.010)	-0.004 (0.010)
is_com	0.417*** (0.071)	0.422*** (0.071)	0.416*** (0.072)
has_hyphen	0.063 (0.237)	0.062 (0.236)	0.065 (0.237)
has_digit	-0.076 (0.079)	-0.075 (0.080)	-0.077 (0.079)
token_count	-0.125 (0.154)	-0.126 (0.154)	-0.126 (0.155)
exact_match	-0.036 (0.435)	0.120 (0.445)	-0.086 (0.438)
dist_sq		-1.453 (1.221)	
dist_log			-1.799*** (0.453)
Num.Obs.	294	294	294
R2	0.254	0.257	0.252
R2 Adj.	0.220	0.219	0.218
R2 Within	0.189	0.191	0.187
R2 Within Adj.	0.165	0.165	0.163
AIC	462.6	463.7	463.4
BIC	514.2	518.9	514.9
RMSE	0.51	0.51	0.51
Std.Errors	Heteroskedasticity-robust	Heteroskedasticity-robust	Heteroskedasticity-robust
FE: cluster_keyword	X	X	X

+p < 0.1, \*p < 0.05, \*\*p < 0.01, \*\*\*p < 0.001

- Cairncross, F. (1997). *The death of distance: How the communications revolution is changing our lives*. Boston, MA: Harvard Business School Press.
- Ellison, G., & Ellison, S. F. (2009). Search, obfuscation, and price elasticities on the internet. *Econometrica*, 77(2), 427–452.
- Gaspar, J., & Glaeser, E. L. (1998). Information technology and the future of cities. *Journal of Urban Economics*, 43(1), 136–156.
- Lindenthal, T. (2014). Valuable words: The price dynamics of internet domain names. *Journal of the Association for Information Science and Technology*, 65(5), 869–881.
- Lindenthal, T. (2018). Monocentric cyberspace: The primary market for internet domain names. *Journal of Real Estate Finance and Economics*, 57, 152–166.
- Lucas, R. E., Jr., & Rossi-Hansberg, E. (2002). On the internal structure of cities. *Econometrica*, 70(4), 1445–1476.
- Mills, E. S. (1972). *Studies in the structure of the urban economy*. Baltimore, MD: Johns Hopkins University Press.
- Muth, R. F. (1969). *Cities and housing: The spatial pattern of urban residential land use*. Chicago, IL: University of Chicago Press.
- Pascal, A. (1987). The vanishing city. *Urban Studies*, 24(6), 597–603.
- The Economist Online. (2010, October). Sex. sells. *The Economist*. Retrieved from [http://www.economist.com/blogs/dailychart/2010/10/domain-name\\_prices](http://www.economist.com/blogs/dailychart/2010/10/domain-name_prices)
- Tranos, E., & Nijkamp, P. (2013). The death of distance revisited: Cyber-place, physical and relational proximities. *Journal of Regional Science*, 53(5), 855–873.

TABLE IX

## PART II ROBUSTNESS: HETEROGENEITY BY TLD, AGE, AND LENGTH

	.com only	Non-.com	Younger domains	Older domains	Shorter domains	Longer domains
distance_used	-1.486*** (0.401)	-1.360+ (0.687)	-1.570*** (0.447)	-0.984+ (0.542)	-1.118** (0.374)	-2.131* (0.849)
length	0.001 (0.014)	0.023 (0.030)	0.016 (0.020)	-0.004 (0.016)	-0.011 (0.022)	0.046+ (0.027)
age_years	-0.005 (0.012)	0.000 (0.018)	0.043 (0.078)	-0.015 (0.018)	-0.005 (0.016)	0.004 (0.015)
has_hyphen	0.003 (0.307)	0.198 (0.413)	-0.009 (0.347)	0.139 (0.329)	-0.770*** (0.225)	0.263 (0.467)
has_digit	-0.123 (0.088)	0.025 (0.176)	-0.129 (0.097)	-0.061 (0.117)	-0.166 (0.103)	0.035 (0.127)
token_count	-0.078 (0.200)	-0.235 (0.276)	-0.145 (0.226)	-0.128 (0.218)	0.681*** (0.139)	-0.271 (0.259)
exact_match	-0.148 (0.403)			-0.024 (0.468)	-0.116 (0.517)	
is_com			0.367*** (0.108)	0.490*** (0.096)	0.483*** (0.088)	0.307** (0.111)
Num.Obs.	217	77	147	147	179	115
R2	0.192	0.139	0.284	0.272	0.275	0.328
R2 Adj.	0.145	-0.007	0.220	0.201	0.218	0.249
R2 Within	0.095	0.070	0.189	0.219	0.238	0.172
R2 Within Adj.	0.064	-0.015	0.146	0.172	0.201	0.115
AIC	340.5	133.7	248.5	229.6	283.3	188.0
BIC	384.4	161.8	287.4	271.4	327.9	223.7
RMSE	0.50	0.49	0.52	0.48	0.49	0.49
Std.Errors	Heteroskedasticity-robust	Heteroskedasticity-robust	Heteroskedasticity-robust	Heteroskedasticity-robust	Heteroskedasticity-robust	Heteroskedasticity-robust
FE: cluster_keyword	X	X	X	X	X	X

+ $p < 0.1$ , \* $p < 0.05$ , \*\* $p < 0.01$ , \*\*\* $p < 0.001$

- Greenstone, M., Hornbeck, R., & Moretti, E. (2010). Identifying agglomeration spillovers: Evidence from winners and losers of large plant openings. *Journal of Political Economy*, 118(3), 536–598.
- Moretti, E. (2021). The effect of high-tech clusters on the productivity of top inventors. *American Economic Review*, 111(10), 3328–3375. []
- Garg, V. (2022). Why are cosine similarities of text embeddings almost always positive? *Medium*, available at: <https://vaibhavgarg1982.medium.com/why-are-cosine-similarities-of-text-embeddings-almost-always-positive-6bd31eae4d5>. []
- Enevoldsen, K., Chung, I., Kerboua, I., Kardos, M., Mathur, A., Stap, D., Gala, J., Siblini, W., Krzemiński, D., Winata, G. I., Sturua, S., Utpala, S., Ciancone, M., Schaeffer, M., Sequeira, G., Misra, D., Dhakal, S., Ryström, J., Solomatin, R., Çağatan, Ö., et al. (2025). MMTEB: Massive Multilingual Text Embedding Benchmark. *arXiv preprint arXiv:2502.13595*. []