Prime locations*

Gabriel M. Ahlfeldt† Thilo N. H. Albers‡ Kristian Behrens§

December 9, 2024

Abstract

We develop a method for delineating prime locations—ultra-dense clusters of economic activity—within cities and apply it to comprehensive establishment data covering all US metropolitan areas. We further use big data to extend the analysis to a global context where administrative data are sparse. Many cities in the US and around the world are dominated by a small number of prime locations that can concentrate up to one-half of tradable services jobs on tiny shares of land and are dominant centers of economic gravity. In cities with fewer prime locations, a larger tradable services share concentrates in them.

Keywords. Internal city structure; employment density; clustering; tradable services; point-pattern analysis.

JEL classification. R38, R52, R58.

*We gratefully acknowledge CEP, SERC, and the LSE research committee for funding. We are indebted to Christian Hilber and Felix Weinhardt without whose input this project would not have come about. We thank the editor, Gabriel Chodorow-Reich, and two anonymous reviewers for many useful comments. We also thank participants at numerous seminars and conferences, and especially Nate Baum-Snow, Victor Couture, Felipe Carozzi, Paul Cheshire, Gabriel Felbermayr, Steve Gibbons, Laurent Gobillon, Stephan Heblich, Vernon Henderson, Henry Overman, Olmo Silva, Daniel Sturm, Jens Südekum, Nick Tsivanidis, and Niko Wolf for comments and suggestions. Further thanks are due to Marco Gonzalez-Navarro and Matt Turner for sharing their subway data with us at an early stage of the project; and Jason Barr and Remi Jedwab for sharing their height-gap data with us. Katia Del-Frari, Peter Jeffrey, Kalle Kappner, Stephen Law, Sascha Moebius, Alice Neht, Tiantong Song, Timo Stieglitz, and Sevrin Waights provided excellent research assistance. Behrens gratefully acknowledges financial support from SSHRC (Insight Grant #435-2021-1017). Ahlfeldt & Albers gratefully acknowledge the support of the Deutsche Forschungsgemeinschaft (CRC TRR 190–project #280092119). The usual disclaimer applies.

†Humboldt University Berlin; LSE, CEPR, CESifo, and CEP: g.ahlfeldt@hu-berlin.de.
‡Humboldt University Berlin: thilo.nils.hendrik.albers@hu-berlin.de.
§Université du Québec à Montréal; and CEPR: behrens.kristian@uqam.ca.

1 Introduction

The workhorse urban model remains that of the monocentric city. It features a single employment centre, the central business district (CBD), that serves as an anchor for city structure. In particular, the CBD is surrounded by a residential doughnut within which population density, building heights, and land- and real-estate prices decline towards the outskirts. Although the monocentric model performs well in describing first-order features of urban spatial structure, it has been increasingly challenged by the existence of polycentric cities with multiple large business centers and smaller edge cities. How important are these large centers and what jobs do they concentrate? How many such centers do cities have and how important are they for city structure? Beyond anecdotal evidence and case studies, we have little systematic knowledge about these questions, especially for cities outside the US.

We engage with these questions in three steps. First, we develop a novel method for delineating *prime locations*—ultra-dense clusters of economic activity—within cities. Intuitively, we identify places with employment densities that are highly unlikely to arise under spatial randomness. Second, we apply our method to comprehensive georeferenced establishment data covering 381 US metropolitan statistical areas (MSAs). We delinate 531 prime locations in those MSAs, thus producing the first comprehensive picture of US urban spatial structure. Third, we use big open data to extend our method—using weights estimated from what we have learned in the US—to a global context where administrative data are sparse. We delineate 286 prime locations in 125 global cities across 35 countries to produce a first picture of urban spatial structure across many countries around the world. Equipped with our findings we can start answering the questions we have set out to investigate.

How important are prime locations and what jobs do they concentrate? The most extreme prime location, Midtown in New York, concentrates 1.7 million jobs within an area of 11 km^2, i.e., a staggering 156 thousand jobs per km^2. More generally, many large US cities and global cities in our world-wide sample concentrate more than one-third of tradable services employment on less than 0.3% of developable land. Just 531 disks with a radius of 20 kilometers—centered on our prime locations—contain more than half of the population and two-thirds of tradable services jobs on a territory of almost 10 million square kilometers.

How many such centers do cities have and how important are they for internal city structure? Nearly 70% of our sampled global cities—and more than 90% of US MSAs—have only a small number, one or two, of dominant employment centers. These prime locations are the nuclei of economic gravity in cities: many markers of economic activity fall off steeply as one moves away from prime locations. These findings may explain why—despite the increasing decentralization of population (Baum-Snow, 2007), employment (Glaeser and Kahn, 2004; Baum-Snow et al., 2017), and land values (McMillen, 1996)—the monocentric model continues to perform remarkably well for many cities around the world (see Liotta et al., 2022; Ahlfeldt and Barr, 2022, for recent estimates of population and height gradients).

We use our delineated prime locations to uncover correlations that offer new insights into the determinants of city configurations. Without providing causal or structural interpretations, we document a hitherto unnoticed negative relationship between the number of prime locations and their joint share at a city's tradable services employment, which points to the importance of agglomeration economies within prime locations. Secondary prime locations also tend to be more specialized, suggesting that localization economies may be a driving force of sub-center formation. Since cities that adopted mass transit earlier tend to be more monocentric, and since monocentric cities tend to offer faster commutes, there appear to be scale economies in transportation related to the geography of prime locations.

Our paper makes two main contributions. First, identifying prime locations as business districts of city-wide importance is a challenging task, particularly in a global sample of cities. As administrative definitions of spatial units differ substantially, we cannot rely on them for identifying business districts (Duranton, 2021). Nor can we rely on existing algorithmic alternatives (e.g., McMillen, 2001), which typically delineate subcenters as positive deviations in employment density from a smooth density surface. They thus identify subcenters that *locally* dominate a pre-defined area rather than *globally* dominate the city, such as CBDs do. To overcome the challenge of identifying only centers that carry significance across the city, we develop a new algorithm that draws on the insights from point pattern-based analysis (e.g., Duranton and Overman, 2005). It tests for significant departures from the counterfactual of spatial randomness, conditional on the city's establishment distribution and on locations being developable. Among the identified clusters, it retains those with abnormally high employment density and aggregates them into prime locations. The appeal of our new algorithm is that it allows to identify business centers with citywide importance without restricting their number a priori, and does so using data that are comparable across cities.

Second, we create two new datasets—one for the US and one for our global cities sample—that map the locations of the dominant centers of economic activity. Those datasets, along with a user-friendly Python implementation of our delineation algorithm, are available from our GitHub toolkit. These data and tools should prove useful to researchers who need good measures of where the centers of economic activity are located. The empirical urban economics literature has traditionally struggled with properly defining the 'city center'. Originating with Holian and Kahn (2012), one strategy has been to use the coordinates returned from entering the city name into Google Earth (e.g., Couture and Handbury, 2020; Chodorow-Reich et al., 2024). Another strategy is to use some points of interest such as the geographic center, the historical center, or the town hall (e.g., Liotta et al., 2022). Our approach improves upon these alternatives because we detect all dominant centers in the city—which is especially important for large cities—and map their spatial extents and economic weights.

Our results contribute to three strands of literature. First, we add to recent work that exploits big data. Previous studies use remote-sensed land cover and lights data to estimate population, economic activity, and city geometry (e.g., Henderson et al., 2012; Donaldson and Storeygard, 2016; Harari, 2020), as well as geo-tagged social-media data to predict subnational GDP (Indaco, 2020), among others. However, these data have neither industry nor employment information. We solve this problem by using industry-specific data scraped from the web or queried from Google Places. Second, we complement the literature that aims to create comparable spatial units to study economic questions (e.g., Rozenfeld et al. 2011; Duranton 2015).[1] Our algorithm delineates business districts that are comparable across cities. Contrary to existing work (e.g., McDonald, 1987; Giuliano and Small, 1991; McMillen and McDonald, 1998; McMillen, 2001; Redfearn, 2007), we identify centers that dominate cities globally rather than locally. This yields economic centers that match more closely the idea of a CBD. Last, by documenting that a large fraction of urban economic activity takes place within a small number of prime locations, we build a bridge between the monocentric city model that assumes perfect spatial concentration of employment (e.g., Brueckner, 1987; Fujita, 1989) and quantitative spatial models that can potentially accommodate complete dispersion (e.g. Ahlfeldt et al., 2015; Tsivanidis, 2019; Heblich et al., 2020).

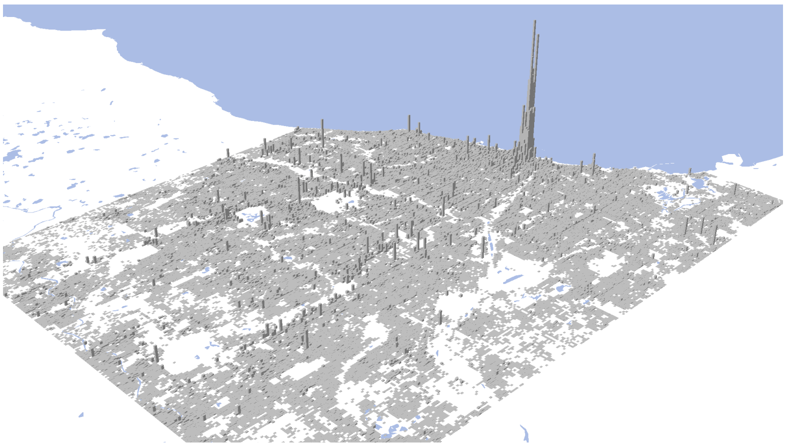
2 The geography of jobs within US metropolitan areas

Using National Economic Time Series (NETS) microdata covering the universe of establishments in all US metropolitan statistical areas (MSA; see Appendix A.1 for a description) and a new delineation algorithm, we document a striking degree of spatial concentration of jobs within prime locations—dense clusters specialized in tradable services.

Figure 1 (panels a, c, and e) visualizes the degree of spatial concentration of economic activity by 250×250 meters grid cells within the three largest US cities. We add a third dimension, namely the height of the bars that is proportionate to total employment in a given grid cell (and hence density). Employment is more concentrated in fewer locations in Chicago and New York than in Los Angeles. Yet, the most salient feature of Figure 1 is that all MSAs have some locations that globally dominate the city in terms of employment. To move beyond visual inspection, we now propose an algorithm to detect and delineate these *prime locations* systematically and show that they are dominated by employment in tradable services industries.

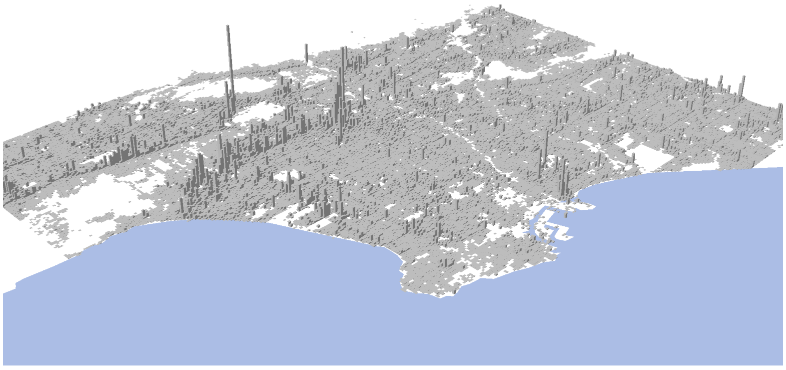
[1] See also the articles in the special issue "Delineating Metropolitan Areas" in the *Journal of Urban Economics* (Duranton and Rosenthal, 2021).

Figure 1: Total employment by grid cell and detected prime locations



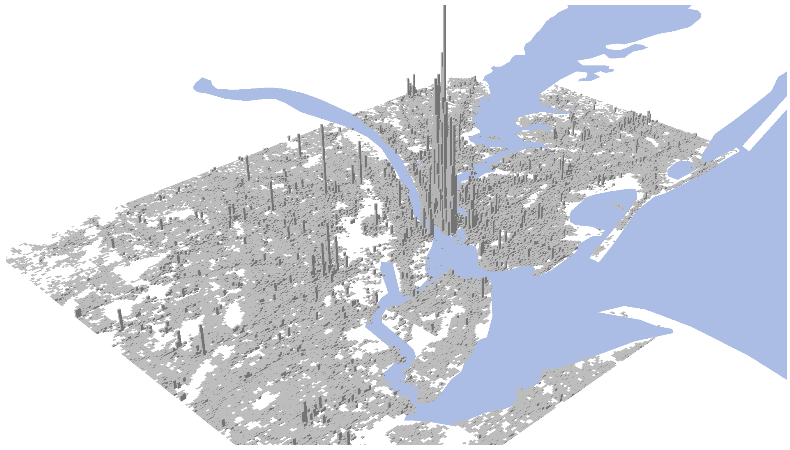
(a) Employment distribution, Chicago

(b) Prime locations, Chicago



(c) Employment distribution, Los Angeles

(d) Prime locations, Los Angeles



(e) Employment distribution, New York

(f) Prime locations, New York

Notes: Establishment-level NETS data for 2012, aggregated to 250×250 meters grid cells. In panels (a), (c), and (e), the extrusion of bars is proportionate to employment density, and the volume of bars is proportionate to employment numbers. Red shaded areas in panels (b), (d), and (f) are prime locations detected using the algorithm described in Section 2.1. Total area of prime locations and employment by prime location and city are documented in Appendix Table B.2.7.

2.1 Detecting and delineating prime locations

We use a novel point pattern-based approach to detect and delineate prime locations. Compared to the literature that separates urban from rural areas, a challenge is that there is no a priori known density threshold for prime locations. We thus must identify boundaries that separate areas of 'unusually high' density from areas of 'ordinary' density.

Our algorithm involves the following steps.[2] For each city, we first randomly draw 100K points uniformly distributed in developable areas within the city. For each point, we compute the employment within a radius of 750 meters. This yields a baseline distribution of employment telling us how likely it is that a randomly drawn point has x workers within a 1.5 kilometer wide disk in the city. We then compute the employment in these disks around each establishment observed in our data (excluding the establishment's own employment) and check where that employment number lies in the city-wide distribution. We call establishments in zones with employment above the 99.5th percentile of the city-wide distribution 'significantly clustered'.[3] We aggregate these significantly clustered establishments to cells, and further aggregate these cells to polygons using the orthogonal convex hull. The resulting areas—clusters of cells with abnormally high employment densities—are what we call prime locations. Figure 1 (panels b, d, and f) maps the identified locations from our algorithm within the three largest MSAs. The locations we identify coincide with the major employment concentrations visible in panels a, c, and e.

Our algorithm has at least four desirable features. First, it uses point data as inputs and is thus not sensitive to spatial aggregation problems.[4] In particular, it avoids the thorny issue of administrative units' comparability, which is especially important when working with cities from multiple countries. Second, it identifies global—rather than local—employment centers within a city's spatial employment distribution. Third, the number of prime locations is a priori unrestricted. Finally, the algorithm is sufficiently flexible to be used in other contexts as the threshold between 'unusually high' and 'ordinary' densities can be easily varied.

2.2 Size and specialization of prime locations

Table 1 summarizes the size and specialization of prime locations in the US. Evidently, prime locations are geographically small but extremely dense clusters. Midtown, NY, is the most concentrated of them. On an area of 11 km^2—equivalent to a circle with a radius of

[2]Appendix B.1 describes the algorithm in pseudo code. Our GitHub toolkit provides an easy-to-use Python version. The faster but more complex version, written in C++, is available in the replication archive.

[3]Too small cutoff values result in larger, but less dense prime locations. Conversely, too high cutoff values weaken the discontinuity at the border of the prime location and exclude dense areas. See Appendix B.1 for a more detailed discussion.

[4]NETS data have been criticised regarding their year-on-year variation and when used at very granular spatial scales (see, e.g., Barnatchez et al. 2021). We show in the appendix (see Figure B.2.3) that our prime locations substantially overlap with the densest areas in MSAs as computed from County Business Patterns data disaggregated to zip code tabulation areas.

about 1.9 kilometers—it hosts a staggering 1.71M jobs. This implies an employment density of close to 156K per km^2. The Loop in Chicago comes second, with 801K jobs and an employment density of 54K per km^2. Lower Manhattan (which includes Wall Street), NY, reaches a higher employment density at close to 110K per km^2, although, at 432K jobs, total employment is smaller owing to its smaller area. To our knowledge, these are the highest employment densities thus far documented in the literature in any context. In other cities, prime locations are smaller but still impressive. Downtown D.C. and Downtown Boston each host more than 500K jobs. In LA, a city known for its polycentric structure, the main prime location hosts more than a quarter of a million jobs at a density of 37K per km^2. Across the 10 largest cities, prime locations host between 7.66% (Miami) and 25.43% (NY) of total employment on 0.14% (Houston) to 0.53% (LA) of developable land. In smaller cities, prime locations are smaller but still concentrate much employment on the head of a pin (see Table B.2.7).

What industries choose to agglomerate at such high densities? Table 1 reveals that the share of jobs in tradable services (defined as NAICS 51–55; see Table A.1.1) within the top-10 prime locations ranges from 27.3% (Downtown D.C.) to 53.2% (Lower Manhatten). This share is generally about twice that outside any prime location in the same city.[5] This specialization suggests that tradable services firms benefit disproportionately from locating within prime locations. Such firms are particularly reliant on knowledge exchange facilitated by face-to-face interactions (e.g., Arzaghi and Henderson, 2008; Baum-Snow et al., 2024). While prime locations nowadays serve as tradable services hubs, many of these locations were originally manufacturing hubs. Indeed, there are still fewer, but stronger prime locations in large cities that had greater manufacturing shares in 1940, suggesting that prime locations underwent a successful long-run process of structural transformation (see Table B.2.1).

2.3 Variation in spatial structure within US cities

Table 1 also shows that prime locations are anchor points of the US economic geography: on average, 1.56M jobs and 2.11M people are located less than 20 kilometers (about the average commuting distance) from the nearest top-10 prime locations in the US. Aggregating to the country level, a staggering 54.41% of the population, 62.42% of the jobs, and 70.13% of tradable services jobs are located within 20km of the centroids of the 531 prime locations we detect. Given the strong gravity of prime locations, it is no surprise that population and housing-unit densities fall by 6.4%, with somewhat steeper gradients around 7-8% for the largest prime locations. We can further corroborate the conventional finding that US suburbs tend to be on average rich as the median household income increases by about 0.5% to 1.5%

[5] Downtown D.C. is the expected exception, due to an exceptionally large public services employment share of more than 50%.

Table 1: Prime locations in US metropolitan statistical areas

MSA	Prime location (PL)	Rank US	Rank MSA	Area (km^2)	Emp. (K)	Density (K/km^2)	TS share (%) PL	TS share (%) outside	Number ≤ 20km emp. (K)	Number ≤ 20km pop. (K)
New York	Midtown	1	1	11.00	1,710	156	47.1	17.9	3,901	5,605
Chicago	The Loop	2	1	14.75	801	54	45.6	17.4	2,036	3,190
Washington	Downtown D.C.	3	1	11.75	671	57	27.8	27.3	1,324	882
Boston	Downtown Boston	4	1	17.31	565	33	45.7	21.6	1,797	2,036
New York	Lower Manhatten	5	2	3.94	432	110	53.2	17.9	2,109	4,125
San Francisco	Financial District	6	1	7.69	421	55	46.3	22.8	991	1,141
Philadelphia	City Center	7	1	12.56	403	32	29.7	19.0	1,096	1,405
Seattle	Seattle	8	1	8.63	278	32	32.9	17.5	534	424
Los Angeles	Downtown Los Angeles	9	1	6.94	255	37	37.0	19.8	1,053	1,833
Atlanta	Downtown Atlanta	10	1	11.56	229	20	32.8	21.3	710	705
Mean, top-10 MSAs				10.61	576	59	39.8	20.3	1,555	2,135
Mean, MSA employment ≥ 1M				3.60	94	20	31.8	18.9	495	667
Mean, MSA employment < 1M				2.08	19	9	19.8	13.6	145	208

Notes: For metropolitan statistical areas (MSAs), we use the first city listed in the official MSA name. For prime locations (PL), we use the colloquial names suggested by ChatGPT. TS share is the share of tradable services (NAICS codes 51-55) employment at total employment (emp.). The last column reports population counts (pop.) computed from 2010 Census blocks and the distance of their centroids to the centroid of the PL. A full-length version of this table for 531 prime locations is provided in our GitHub toolkit.

per kilometer distance from our prime locations (see Table B.2.6).[6]

2.4 Variation in spatial structure between US cities

Our analysis offers new insights into the spatial configuration of US cities. Chicago—the textbook example of a monocentric city—has only one prime location (The Loop), New York two (Midtown and Lower Manhattan), and Los Angeles and Dallas as many as nine. 88.1% of small US cities are monocentric. Among large cities (with 1M jobs or more), only 11.1% are monocentric, whereas 16.7% or duocentric and the remaining 72.2% are polycentric. However, even most polycentric cities have only a limited number of centers: only 9 cities have 5 or more prime locations, and none has 10 or more. New York has the largest and most specialized prime locations (see Table 1), with 25.43% of city employment and 48.1% of city-wide tradable services (TS) employment concentrated in Midtown and Lower Manhattan. On the opposite, Miami has only 27.66% of its city employment and 11.66% of TS employment located in its 8 prime locations. On average, large cities have 13.27% of city-wide employment and 21.22% of city-wide TS employment concentrated on 0.28% of their developable area. The corresponding figures for smaller cities (less than 1M employment) are 0.76% of city-wide employment, 14.85% of city-wide TS employment, and 0.1% of developable area.

Panel (a) of Figure 2 reveals that there is a positive association between city size and the number of prime locations, consistent with prior evidence on larger US cities having more sub-centers (McMillen and Smith, 2003). For MSAs with more than 1M jobs (large

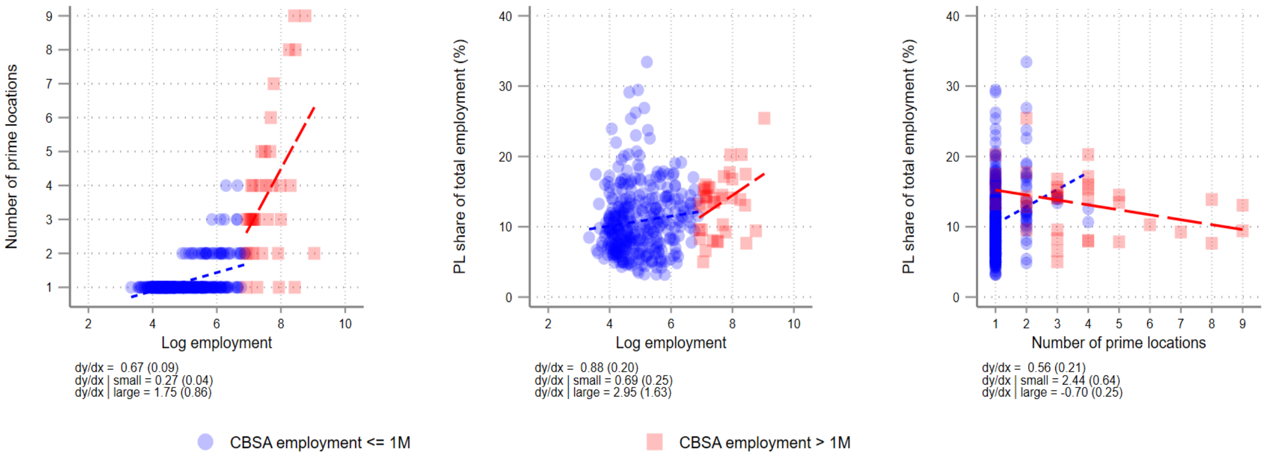
[6]Results for some individual cities, such as New York, show a negative income gradient. Glaeser et al. (2008) document some cross-city heterogeneity in income gradients and show that those of 'old' cities (NY, Chicago, Philadelphia) differ from those of 'new' cities (Atlanta, Phoenix, LA).

cities), a doubling in city size is associated with $1.75 \times \ln 2 \approx 1.2$ additional prime locations (left exhibit; red long-dashed line). We add the novel observation that there is also a positive association between city size and a *city's share of jobs in prime locations*. Among large cities, a doubling of city size is associated with an increase in prime locations' share of total employment by $0.69 \times \ln 2 \approx 0.48$ percentage points (middle exhibit; red long-dashed line), about 15% of the mean share (see Table B.2.7). This increase is not mechanical in the sense that more prime locations means more jobs in prime locations. To the contrary, across large cities, *the share of prime locations at total jobs falls in the number of prime locations* (right exhibit; red long-dashed line). One additional prime location decreases the share of jobs in all prime locations by 0.7 percentage points on average. To pick one telling example, total employment in LA's nine prime locations jointly accounts for a much smaller share in citywide total employment (less than 10%) than the two prime locations in New York (about 25%).[7] This novel stylized fact suggests a sub-additivity property, where a single prime location exerts a stronger pull than multiple smaller ones. This phenomenon can be attributed to localized agglomeration economies: dividing workers across multiple locations may reduce productivity due to lost agglomeration benefits. We think this finding should spur additional research on the determinants and the efficiency of internal city structure, an area where empirical research lags behind theory (Fujita and Ogawa, 1982; Lucas and Rossi-Hansberg, 2002).

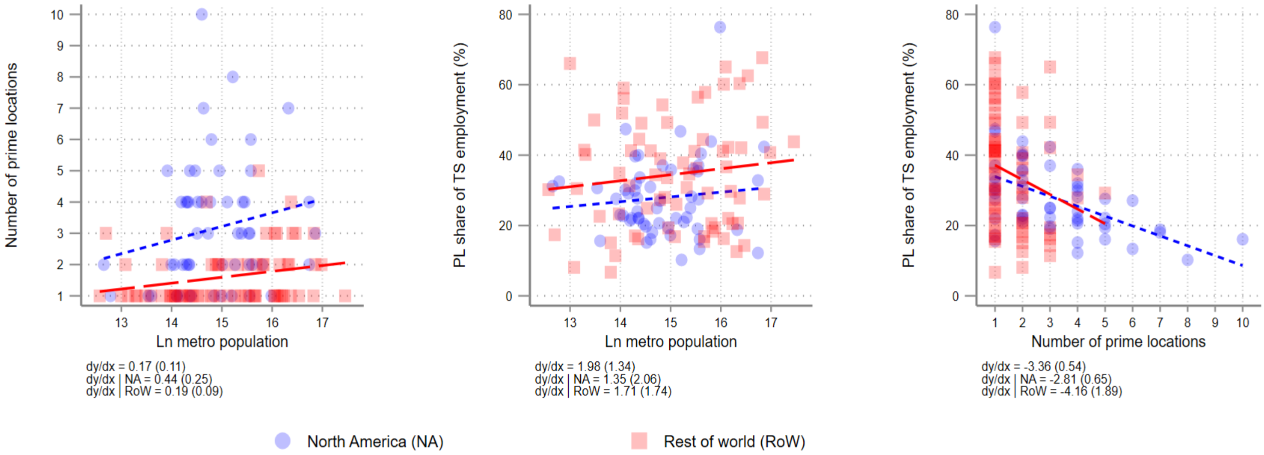
The rightmost exhibit in panel (a) of Figure 2 shows that sub-additivity is especially a phenomenon of large cities: within small US cities the relationship is actually positive (blue short-dashed line). The notion that agglomeration economies are especially important in large cities is substantiated by prime locations being more specialized when there are fewer prime locations in large cities (in small cities, it is the opposite). Within large and small cities, lower-ranking prime locations (in terms of employment size), tend to be more specialized, suggesting that one reason why secondary prime locations emerge is to capitalize on localization economies (see Table B.2.4). In this context, it is worth noting that horizontal land use regulations may be a deterrent to the formation of secondary prime locations as we find a negative correlation between an MSA's number of prime locations and the Wharton Residential Land Use Regulatory Index (Gyourko et al., 2021) (see Table B.2.2). Vertical land use regulation may push in the opposite direction as we find cities that constrain demand for height (Barr and Jedwab, 2023) to have more prime locations that are jointly less important (see Table B.2.3).

[7] See Table B.2.7. Our GitHub toolkit provides a classification of the degree of polycentricity and agglomeration of jobs in prime locations for all 381 MSAs.

Figure 2: Number and importance of prime locations



(a) US MSAs (all employment)



(b) Global cities sample (tradable services employment)

Notes: Each dot represents one city. Shares refer to the collective share of all prime locations within a city at employment in the entire city. Panel (a) presents results for 381 US MSAs. Panel (b) presents results for 125 global cities. In panel (b), North American (NA) cities include 39 US and 9 Canadian cities. Rest of world (RoW) cities comprise 77 cities in 33 countries.

3 Prime locations in 125 global cities

Our findings for the US suggest that the sub-additivity of prime locations is most relevant for the large prime locations of 'global cities', such as New York or Chicago. We now extend the scope of our analysis to 125 global cities. Doing so will allow us to show that our key insights generalize beyond US MSAs.

3.1 Extending the analysis to a global sample of cities

We select a sample of 125 global cities in 35 countries. To be included in the sample, a city needs to show 'globalized economic activity' as measured by a minimum number of grade-A office buildings held by Real Estate Investment Trusts (SNL-S&P) and of Starbucks franchises (available as an open source dataset). Our data-based and automated selection delivers a final set of cities that shows a large overlap with existing global city lists (PricewaterhouseCoopers, 2009; Trujillo and Parilla, 2016), but excludes many large cities in the developing world, especially in Africa and India.[8] This section discusses how we apply big data techniques to create an establishment-level dataset of tradable services employment for these cities and delineate their prime locations.

Global establishment-level employment data. The methodology developed in Section 2.1 requires geocoded establishment-level employment data, which are not available for most cities around the world. As shown above, however, prime locations are highly specialized in tradable services. To gather geocoded data on these types of establishments, we design a strategy that extracts both *local tradable services establishments* as well as *global industry leaders*. For the local establishments, we run queries within Google Places API for the following establishment types, which we index by $s \in S$: accounting firms, consultancies, insurance, investment banks, and law firms. We augment this dataset by scraping establishment data from the respective global leaders' websites in these sectors (such as the top-4 accounting firms, the top-5 consultancies, and the top-10 global law firms). Additionally, we locate central banks, stock exchanges, and the headquarters of all tradable services companies listed in the respective country's leading stock market index (e.g., the S&P 500 for US cities). In doing so, we gather more than 100K 'big data' establishments across $S = 13$ categories. Appendix A.2 provides an extensive documentation of these data.

Although the establishments are geocoded, we have no information on employment. To overcome this limitation, we estimate the relationship between the density of big data establishments and the density of employment using the 39 large US MSAs in our global cities sample that overlap with the NETS data. We begin by drawing 100K randomly distributed points indexed by $d \in D$ within the developed area of each of our 39 US cities that we index by $m \in M$. Around each point, we create a 750 meters buffer and compute the total employment in tradable services observed in the NETS data, $E_{d,m}^{TS}$, as well as S count measures of the big data establishments within that buffer. To estimate employment weights, we then run the following Non-Negative Least Squares regression:[9]

$$E_{d,m}^{TS} = \mathbf{X'_{d,m}} \mathbf{b_{s,m}} + \varepsilon_{d,m}, \tag{1}$$

where $\mathbf{X_{d,m}}$ is our $(D \times M) \times S$ matrix of big data establishment counts and $\mathbf{b_s}$ is the associated $1 \times (S \times M)$ vector of city-specific employment weights. We omit city-specific in-

[8] Appendix A.2 reports details on city selection and delineation. See the notes to Table 2 for a complete list of cities by world region.

[9] The NNLS algorithm by Lawson and Hanson (1974) is used to estimate linear models where coefficients are constrained to be non-negative. The algorithm iteratively adjusts coefficients by moving variables in and out of an active set, optimizing the fit while maintaining non-negativity constraints.

tercepts to be consistent with the prediction within the global sample of cities where we will naturally obtain zero employment if there are no big data establishments. Since establishment size may vary by city size, we parametrize city-type employment weights as $\mathbf{b_{s,m}} = \exp\left(b_s^0 + b_s^1 \ln L_m\right) + \epsilon_{m_s}$, where L_m is metro area population, ϵ_{m_s} is a residual, and $\{b_s^0, b_s^1\}$ are parameters that we estimate using a Poisson maximum likelihood estimator.

In keeping with intuition, we find much larger employment weights for the scraped establishments that consist of multinational industry leaders than for the average establishment listed in Google Places, especially in large cities (see Appendix B.3.1 for the estimated weights). This is consistent with the microgeographic sorting of high-quality establishments into high-quality locations and peer groups, as documented by Baum-Snow et al. (2024). In the last step, we assign the estimated employment weights, $\widehat{\mathbf{b}}_{\mathbf{s,m}} = \exp\left(\widehat{b}_s^0 + \widehat{b}_s^1 \ln L_m\right)$, to the big data establishments.

Identification of prime locations in 125 cities. We identify 286 prime locations across 125 cities (the full set of cities and number of prime locations are given in the notes to Table 2). Before we further explore the variation in the geography of prime locations in our global cities sample, we provide additional evidence for the validity of the approach we have developed.

Validation. Our approach towards detecting and delineating prime locations in sparse administrative data environments using big open data rests on several assumptions. In the following, we summarize these and discuss how our validation exercises speak to them (Appendix B.3.3 reports further details). First, establishments in tradable services alone are sufficient to accurately delineate prime locations. To evaluate this assumption, we replicate the delineation of prime locations using solely tradable services establishments from the NETS data. Second, we claim that the employment-weighted big data establishments are a good proxy for the spatial distribution of tradable services jobs. To evaluate this assumption, we replicate the delineation using establishments from the NETS data restricted to the sectors that correspond to the keywords used in our big data approach. Third, we claim that the employment-weighted big data establishments are a good proxy for the spatial distribution of tradable services jobs. To evaluate this assumption, we subject our entire process of collecting big data and assigning employment weights to a rigorous overidentification test. We focus on the 39 US MSAs for which we have NETs data and collected big data establishments and split the sample of MSAs into two mutually exclusive halves by alphabetical order. In two separate exercises, we use one of the sub-samples to predict employment weights and the other to construct employment weights for the big data establishments which are then input into the delineation of prime locations. In each case, we compare the delineated prime locations—based on alternative employment measures—to the prime locations delineated in Section 2.1. Reassuringly, conditional on a grid cell being part of a prime location delineated using any proxy measure, the probability of it being part of a prime location identified with the most comprehensive data ranges from 71.2% to 82.4%. Naturally, we do not have micro-geographic employment data for all cities in our sample—the very reason for developing our approach. Nevertheless, we show that there is a high correlation, significant at 1%, between the predicted employment in tradable services and other typical markers of prime locations: grade-A office stock, coworking spaces, and Starbucks franchises. To summarize, we are confident that our approach picks up well the prime locations across the cities in our global sample.

3.2 Variation in spatial structure within world cities

The average city in our global sample has 2.3 prime locations, less than the 3.79 prime locations in large US cities.[10] The mean geographic area of a prime location is 2.6 km^2 in the global sample, somewhat smaller than the 3.6 km^2 documented for large US cities. The prime location share at tradable services employment, at 31.6% on average, however, is greater than that of the average large city in the US (21.2%). Finally, as in the 381 US cities, prime locations are extremely dense clusters. On average, they only occupy 0.26% of a city's area around the globe (compared to 0.28% for the large US cities).

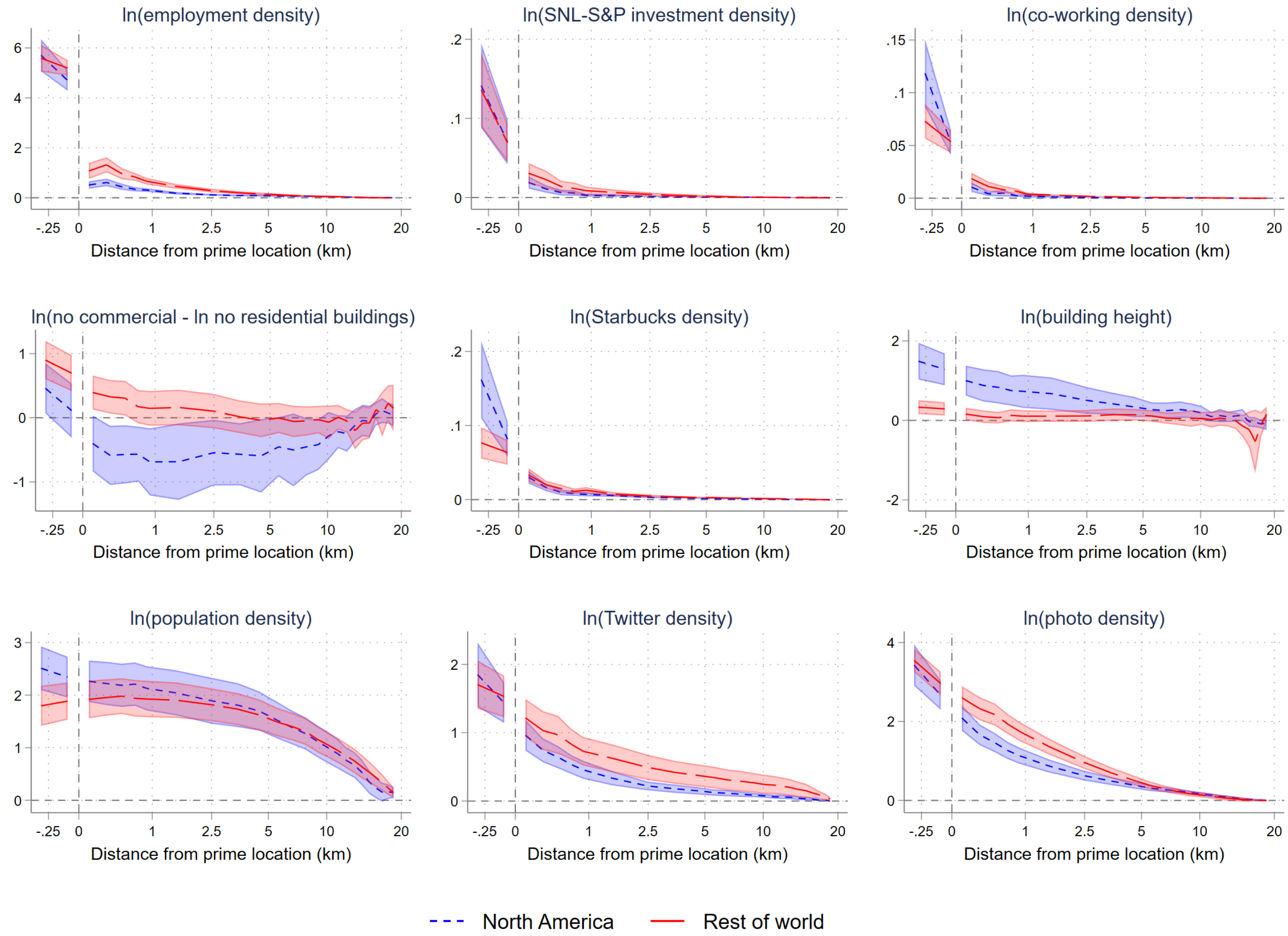
To show how strongly these prime locations anchor economic activity, we gather data on building heights and counts, population, social media activity, grade-A office stock, georeferenced Starbucks franchises, and georeferenced coworking spaces to estimate distance gradients, the nuclei of which are our prime locations. We present non-parametric estimates of these distance gradients separately for North America and the rest of the world in Figure 3.

The first row plots the density gradients for the predicted tradable services employment, investments into grade-A office buildings, and coworking spaces. The gradients show stark discontinuities at the prime location borders: slopes flatten and densities fall sharply. The predicted tradable services employment density falls by up to $e^{-4} - 1 = 98\%$ at the border, on average. Gradients for grade-A office stock investments (SNL-S&P) and coworking spaces (scraped from Regus and WeWork) exhibit a strong fall too, and a rather sharp discontinuity at the prime locations' borders. These two variables capture how the market evaluates the long-run (investments) and short- and medium-run demand (coworking) for office space. They have not been used to delineate prime locations.

The second row of Figure 3 shows that the share of commercial buildings (Emporis) drops quickly as we leave prime locations, as does the density of Starbucks franchises. The latter is an iconic visible indicator of business districts in the US, hence it is no surprise that Starbucks' density peaks at a higher value for North American cities. Height (Emporis) gradients fall off, too, even more so for North American cities where height constraints

[10] Appendix B.3.2 reports corresponding summary statistics.

Figure 3: Prime location distance gradients



Notes: The gradients are averaged across up to 286 prime locations in 125 global cities, depending on data availability for the particular outcome. The dashed line marks the border of the prime location. Negative distances on the x-axis (left of the dashed line) describe the distance to the border of points within the prime location, positive distances (right of the dashed line) describe the distance to the border of points lying outside of the prime location. Underlying each figure is a grid-level regression of the outcome against 27 distance bins (bin size for -.4-1km: 200m; for 1km-20km: 1km; reference category is the bin covering 20-21km) and a city fixed effect. Shaded areas represent 90% confidence intervals of the respective point estimates. The underlying employment data is our prediction based on the 'big data'. For a description of sources for the other outcomes, see Appendix A.2.2.2.

within business districts tend to be less binding and, perhaps as a result, the degree of land-use segregation is somewhat greater.

One may worry that our prime locations pick up urban density in general, not primarily density related to business activity. The final row of gradients in Figure 3 provides falsification tests. In particular, gradients for variables relating to urban density more generally exhibit a different curvature from those in the first two rows. First, the population density (Gridded Population of the World) remains fairly flat at a high level (or even increases in RoW cities) outside of the prime location until a distance of around 5 kilometers. Second, the density of tweets (Gnip) drops as we move away from the prime location. However, there is no strong visible discontinuity in either levels or slopes at the border. The same holds true for the density of social media photos (Flikr and Picasa). Hence, our approach picks up the dominant business centers of cities and not just urban density per se.

To summarize, we find prime location distance gradients that are not only consistent with the predictions of the standard urban model, but also qualitatively and, mostly, quantitatively similar for cities within and outside North America.[11]

3.3 Variation in spatial structure between world cities

Column (1) in Table 2 points to the existence of substantial cross-continental heterogeneity in urban structure. North American cities, serving as the reference category, have three prime locations on average. Compared to those, cities in Australia, Europe, and South America have fewer (panel A) that jointly exhibit a similar or even higher collective share in tradable services employment (panel B). Australian cities show the strongest degree of concentration with 1.3 prime locations on average, which house up to 36.6% of the citywide tradable services employment.

Column (2) suggests that neither the size of cities per se nor first-nature features such as waterbodies and terrain ruggedness are sufficient to rationalize these differences across continents. However, they are important correlates of how many centers a city has. First, a doubling of metro area population is associated with $0.25 \times \ln 2 \approx 0.17$ additional prime locations, on average. There is, however, no significant relationship with prime locations' employment share. Hence, prime locations in larger cities tend to be smaller in relative terms. Second, mountains and water bodies represent barriers that may constrain city shape. Access to water may also represent an amenity that strengthens an existing prime location, making it difficult for a second prime location to emerge. Indeed, we find weak evidence that a larger share of water anchors the geography of jobs and is associated with a larger employment share in fewer prime locations. Steep slopes are associated with more prime locations, but there is no relationship with prime locations' employment shares.

Mass transit systems allow residents to access more jobs at lower congestion costs. Narratives about urban form often cite this transportation technology, particularly if established early, as anchoring central business districts.[12] Intuitively, hub-and-spoke mass transit systems favor highly concentrated urban structures since they maximize labor supply at their central nodes. Columns (3) and (4) explore this argument. Cities that just introduced a subway system have, all else equal, 0.35 more prime locations, on average, but the effect is not statistically significant. This number drops by 0.011 for each year that has passed since the subway opening and this marginal effect is statistically significant. A city that introduced a subway system a century ago has $100 \times 0.011 - 0.35 = 0.75$ fewer prime locations than an otherwise comparable city. At the same time, the employment share of its prime locations is larger by $100 \times 0.129 - 6.047 = 6.853$ percentage points. Notably, the gap in the

[11] Likewise, gradients are similar for monocentric, duocentric, and polycentric cities (see Appendix B.4).

[12] See, for example, Glaeser (2012, p. 141): *"Transportation technologies shape cities, and Midtown Manhattan was built around two great rail stations that could carry in oceans of people."*

Table 2: Determinants of the geography of prime locations

	(1)		(2)		(3)		(4)	
	Panel A: Number of prime locations							
Africa & Middle East	-1.145	(0.556)	-1.087	(0.624)	-1.345	(0.675)	-1.355	(0.690)
Asia	-1.145	(0.364)	-1.497	(0.398)	-1.867	(0.412)	-1.826	(0.411)
Australia	-1.812	(0.392)	-1.736	(0.366)	-1.914	(0.383)	-1.940	(0.390)
Europe	-1.793	(0.297)	-1.682	(0.299)	-1.297	(0.286)	-1.272	(0.289)
South America	-1.745	(0.356)	-2.197	(0.427)	-2.441	(0.498)	-2.475	(0.502)
Ln metro population			0.249	(0.134)	0.477	(0.178)	0.510	(0.189)
Share water (%)			-0.020	(0.012)	-0.017	(0.012)	-0.017	(0.012)
Share steep slope (%)			0.025	(0.013)	0.020	(0.014)	0.021	(0.014)
Subway					0.350	(0.451)	0.335	(0.457)
Years since subway opening					-0.011	(0.003)	-0.010	(0.004)
# subway stations							-0.002	(0.001)
Constant	3.145	(0.276)	3.191	(0.266)	3.546	(0.296)	3.552	(0.297)
R^2	.228		.282		.322		.324	
	Panel B: Prime locations' share at employment (%)							
Africa & Middle East	1.177	(5.674)	-0.723	(4.592)	1.195	(4.790)	1.572	(4.524)
Asia	9.644	(3.754)	10.472	(4.430)	14.974	(4.194)	13.537	(4.199)
Australia	8.723	(2.985)	8.504	(3.847)	10.012	(3.639)	10.938	(3.655)
Europe	3.456	(3.103)	3.828	(2.930)	0.023	(3.568)	-0.854	(3.610)
South America	21.374	(6.488)	18.810	(6.488)	21.674	(5.662)	22.861	(5.387)
Ln metro population			-0.503	(1.491)	-2.603	(1.622)	-3.743	(1.602)
Share water (%)			0.263	(0.106)	0.233	(0.106)	0.256	(0.105)
Share steep slope (%)			0.032	(0.107)	0.089	(0.107)	0.080	(0.097)
Subway					-6.047	(3.010)	-5.528	(3.010)
Years since subway opening					0.129	(0.044)	0.067	(0.050)
# subway stations							0.055	(0.025)
Constant	27.887	(1.536)	27.840	(1.635)	23.703	(1.845)	23.472	(1.792)
R^2	.126		.168		.235		.274	

Notes: The unit of observation is a city. There are 125 observations in all models. Our sample includes the following 125 cities (number of prime locations in parentheses). Africa and Middle East: Cape Town (1), Dubai (4), Durban (2), Johannesburg (3), Kuwait City (1), Riyadh (1), Asia: Bangkok (1), Beijing (3), Chengdu (2), Chongqing (3), Guangzhou (2), Hangzhou (1), Jakarta (1), Kuala Lumpur (5), Nagoya (1), Nanjing (2), Ningbo (1), Osaka (1), Quezon City (4), Seoul (3), Shanghai (2), Singapore (1), Suzhou (2), Taipei City (2), Tianjin (1), Tokyo (1), Wuhan (3), Wuxi (2), Australia: Brisbane (1), Melbourne (1), Sydney (2), Europe: Amsterdam (3), Ankara (1), Antwerp (1), Athens (2), Barcelona (1), Basel (1), Berlin (2), Birmingham (1), Brussels (1), Budapest (1), Dublin (1), Duesseldorf (1), Frankfurt (1), Gothenburg (1), Hamburg (1), Helsingborg (1), Helsinki (1), Istanbul (3), Linköping (1), London (2), Lyon (1), Madrid (1), Malmö (1), Manchester (1), Moscow (2), Munich (1), Paris (1), Reading (3), Rotterdam (2), Stockholm (1), Utrecht (2), Vienna (1), Zurich (1), Örebro (1), North America: Atlanta (4), Austin (4), Boston (2), Calgary (1), Charlotte (2), Chicago (1), Cincinnati (2), Cleveland (1), Columbus (2), Dallas (3), Denver (4), Detroit (4), Edmonton (2), Fort Worth (1), Ft. Lauderdale (6), Guadalajara (3), Hong Kong (2), Houston (2), Indianapolis (4), Kansas City (4), Las Vegas (7), Lexington (2), Los Angeles (7), Mexico City (3), Miami (5), Milwaukee (1), Minneapolis (1), Monterrey (3), Montreal (1), Nashville (5), New York (2), Newark (4), Orlando (1), Ottawa (1), Philadelphia (3), Phoenix (8), Pittsburgh (2), Portland (3), Providence (2), Quebec (1), Riverside (10), Sacramento (5), Saint Louis (4), San Antonio (4), San Diego (5), San Francisco (2), San Jose (5), Seattle (1), Tampa (6), Toronto (2), Vancouver (3), Victoria (1), Virginia Beach (4), Washington (4), Winnipeg (1), South America: Buenos Aires (1), Lima (1), Rio de Janeiro (1), Santiago (2), Sao Paulo (2). In all regressions, the continent indicator variable for North America is the excluded reference category. The employment share is calculated based on our tradable services employment prediction. For subway-related variables, we use and augment data from Gonzalez-Navarro and Turner (2018). Appendix B.4.1 reports the number of prime locations and their employment share by continent. Huber-White heteroscedasticity robust standard errors in parentheses.

number of prime locations between European and North American cities drops by about a quarter, suggesting a role for early subway adoption in explaining why European cities are relatively monocentric. The final column shows that the size of the mass transit network hardly affects the number of prime locations but is weakly associated with an increase in employment share. Hence, the intensive margin result for the importance of prime locations may be driven by the fact that older transit systems are larger and allow to draw in more workers into their central nodes. Controlling for terrain features (water and slopes) and, in particular, subway systems, reverses the positive association between city size and the prime location share of total employment observed in Figure 2. This suggests that subway systems are one explanation for why prime locations in larger cities tend to be more dominant.

The marginal effects for a covariate typically have different signs across the two panels in Table 2. This points to an inverse relationship between the number of prime locations (panel A) in a city and their share at total tradable services employment (panel B). We have previously encountered this sub-additivity property for the case of large US MSAs, and it generalizes to our global cities sample: few cities have high values for both variables. Instead, cities where the largest share of tradable services employment is contained within prime locations are those that have one or few prime locations (among them, Chicago and New York), as shown in panel (b) of Figure 2. The city-wide share of tradabe services concentrated in prime locations falls by some statistically significant 3.36 percentage points for each additional prime location in our global cities sample. The effect is highly robust; even conditional on the controls in Table 2 and country fixed effects, the point estimate is -2.91, statistically significant at the 1% level.

4 Conclusion

Despite decades of decentralization, many cities are dominated by a few large employment centers. Yet, cities also vary significantly in terms of the number and importance of these prime locations. Understanding the origins of these differences represents a rich research agenda. Our findings point to several potential drivers, including gains from specialization, the presence of subway systems, land use regulations, historical industry composition, and natural amenities or geographic barriers. While we document provocative correlations, they remain just that—correlations. Further research is needed to uncover the causal mechanisms and understand how these generate cross-continent heterogeneity in urban spatial structure. We hope that our methodology and data will spur follow-up work to deepen our understanding of the internal structure of cities worldwide.

References

Ahlfeldt, Gabriel M. and Jason Barr, "The economics of skyscrapers: A synthesis," *Journal of Urban*

Economics, 2022, *129*, *103419*.

—, **Stephen J. Redding, Daniel M. Sturm, and Nikolaus Wolf**, "The economics of density: Evidence from the Berlin Wall," *Econometrica*, 2015, *83* (6), 2127–2189.

Arzaghi, Mohammad and J. Vernon Henderson, "Networking off Madison Avenue," *The Review of Economic Studies*, 10 2008, *75* (4), 1011–1038.

Barnatchez, Keith, Leland Dod Crane, and Ryan Decker, "An Assessment of the National Establishment Time Series (Nets) Database," Working Paper 2017-110, Federal Reserve Bank 2021.

Barr, Jason and Remi Jedwab, "Exciting, boring, and nonexistent skylines: Vertical building gaps in global perspective," *Real Estate Economics*, 2023, *51* (6), 1512–1546.

Baum-Snow, Nathaniel, "Did highways cause suburbanization?," *The Quarterly Journal of Economics*, 2007, *122* (2), 775–805.

—, **Loren Brandt, J. Vernon Henderson, Matthew A. Turner, and Qinghua Zhang**, "Roads, railroads, and decentralization of Chinese cities," *The Review of Economics and Statistics*, 2017, *99* (3), 435–448.

—, **Nicolas Gendron-Carrier, and Ronni Pavan**, "Local productivity spillovers," *American Economic Review*, 2024, *114* (4), 1030–1069.

Brueckner, Jan K., "The structure of urban equilibria: A unified treatment of the muth-mills model," *Handbook of Regional and Urban Economics, Chapter 20*, 1 1987, *2*, 821–845.

Chodorow-Reich, Gabriel, Adam Guren, and Timothy McQuade, "The 2000s Housing Cycle With 2020 Hindsight: A Neo-Kindlebergerian View," *Review of Economic Studies*, 2024, *9* (2), 785–816.

Couture, Victor and Jessie Handbury, "Urban revival in America," *Journal of Urban Economics*, 2020, *119*, 103267.

Donaldson, Dave and Adam Storeygard, "The view from above: Applications of satellite data in economics," *Journal of Economic Perspectives*, 2016, *30* (4), 171–98.

Duranton, Gilles, "A proposal to delineate metropolitan areas in Colombia," *Desarrollo y Sociedad*, 2015, (75), 223–264.

—, "Classifying locations and delineating space: An introduction," *Journal of Urban Economics*, 2021, *125*, 103353.

— **and Henry G. Overman**, "Testing for Localization Using Micro-Geographic Data," *The Review of Economic Studies*, 10 2005, *72* (4), 1077–1106.

— **and Stuart S. Rosenthal**, "Special issue on delineation of urban areas," *Journal of Urban Economics*, 2021, *125*, 103352.

Fujita, Masahisa, "Urban economic theory," *Cambridge Books*, 1989.

— **and Hideaki Ogawa**, "Multiple equilibria and structural transition of non-monocentric urban configurations," *Regional Science and Urban Economics*, 1982, *12* (2), 161–196.

Giuliano, Genevieve and Kenneth A. Small, "Subcenters in the Los Angeles region," *Regional Science and Urban Economics*, 1991, *21* (2), 163–182.

Glaeser, Edward L., *Triumph of the city : How our greatest invention makes us richer, smarter, greener, healthier and happier*, London: Pan, 2012.

Glaeser, Edward L and Matthew E. Kahn, "Sprawl and urban growth," in "Handbook of regional and urban economics," Vol. 4, Elsevier, 2004, pp. 2481–2527.

Glaeser, Edward L., Matthew E. Kahn, and Jordan Rappaport, "Why do the poor live in cities? The role of public transportation," *Journal of Urban Economics*, 2008, *63* (1), 1–24.

Gonzalez-Navarro, Marco and Matthew A. Turner, "Subways and urban growth: Evidence from earth," *Journal of Urban Economics*, 11 2018, *108*, 85–106.

Gyourko, Joseph, Jonathan S. Hartley, and Jacob Krimmel, "The local residential land use regulatory environment across US housing markets: Evidence from a new Wharton index," *Journal of*

Urban Economics, 2021, *124*, 103337.

Harari, Mariaflavia, "Cities in bad shape: Urban geometry in India," *American Economic Review*, 2020, *110* (8), 2377–2421.

Heblich, Stephan, Stephen J. Redding, and Daniel M. Sturm, "The making of the modern metropolis: Evidence from London," *The Quarterly Journal of Economics*, 2020, *135* (4), 2059–2133.

Henderson, J. Vernon, Adam Storeygard, and David N. Weil, "Measuring economic growth from outer space," *American Economic Review*, 2012, *102* (2), 994–1028.

Holian, Matthew J. and Matthew E. Kahn, "The impact of center city economic and cultural vibrancy on greenhouse gas emissions from transportation," Technical Report, Mineta Transportation Institute 2012.

Indaco, Agustín, "From Twitter to GDP: Estimating economic activity from social media," *Regional Science and Urban Economics*, 2020, *85*, 103591.

Lawson, Charles L. and Richard J. Hanson, *Solving Least Squares Problems*, Englewood Cliffs, NJ: Prentice-Hall, 1974.

Liotta, Charlotte, Vincent Viguié, and Quentin Lepetit, "Testing the monocentric standard urban model in a global sample of cities," *Regional Science and Urban Economics*, 2022, *97*, 103832.

Lucas, Robert E. Jr. and Esteban Rossi-Hansberg, "On the internal structure of cities," *Econometrica*, 2002, *70* (4), 1445–1476.

McDonald, John F, "The identification of urban employment subcenters," *Journal of Urban Economics*, 1987, *21* (2), 242–258.

McMillen, Daniel P., "One Hundred Fifty Years of Land Values in Chicago: A Nonparametric Approach," *Journal of Urban Economics*, 1996, *40* (1), 100–124.

—, "Nonparametric employment subcenter identification," *Journal of Urban Economics*, 2001, *50* (3), 448–473.

— **and John F. McDonald**, "Suburban subcenters and employment density in metropolitan Chicago," *Journal of Urban Economics*, 1998, *43* (2), 157–180.

— **and Stefani C. Smith**, "The number of subcenters in large urban areas," *Journal of Urban Economics*, 5 2003, *53* (3), 321–338.

PricewaterhouseCoopers, " UK Economic Outlook (November 2009) – Which are the largest city economies in the world and how might this change by 2025?," 2009.

Redfearn, Christian L., "The topography of metropolitan employment: Identifying centers of employment in a polycentric urban area," *Journal of Urban Economics*, 2007, *61* (3), 519–541.

Rozenfeld, Hernán D., Diego Rybski, Xavier Gabaix, and Hernán A. Makse, "The area and population of cities: New insights from a different perspective on cities," *American Economic Review*, August 2011, *101* (5), 2205–25.

Trujillo, Jesus Leal and Joseph Parilla, *Redefining global cities*, The Brookings Institution, 2016.

Tsivanidis, Nick, "Evaluating the Impact of Urban Transit Infrastructure: Evidence from Bogotá's TransMilenio," *Working Paper*, 2019.

Online Appendix for "Prime locations"

Gabriel M. Ahlfeldt Thilo N.H. Albers Kristian Behrens

Conditionally accepted, **American Economic Review: Insights**

A Data

A.1 US data

National Establishment Time Series Our main data source of micro-geographic employment and establishment data is the (NETS). The NETS data are a consistent time-series version of the Dun & Bradstreet (D&B) archival data provided by Walls & Associates.[1] These data contain a quasi-exhaustive picture of all establishments that figure in the D&B databases and have a DUNS number. We have establishment location (latitude and longitude), employment, and primary industry (NAICS 6-digit) codes. More than 95% of establishments and employment are geocoded to the most precise (block-face) level. To stay close to the years for which we have the remaining data for our analysis, we use the 2012 vintage of the NETS data for all our computations.

For the purpose of this study, we aggregate the universe of 2-digit NAICS codes into five larger categories (Table A.1.1) to differentiate the urban economy by sector. Table A.1.2 provides the corresponding breakdown by city and sector for both tradable services and all sectors.

County business patterns For the delineation of city grids and validation checks, we use alternative employment data from the County Business Patterns (CBP), published by the U.S. Census Bureau. We use the 2015 edition at the zip code tabulation area level, which is the finest spatial scale at which these data are available.

Delineation of city grids We proceed in the following steps. First, we identify the zip code with the greatest employment density within five kilometers around its centroid within each MSA according to CBP data. We then put a 70×70km grid (with 280×280 cells of 250×250 meters) around this centroid. We intersect the resulting grid with the official MSA deliniation and only keep the intersecting cells. We then use land cover and digital elevation model (DEM) data to flag undevelopable cells as those the are a majority of water or that have steep slope. For a small number of MSAs, we deviate from this procedure slightly. First, in a very few cases, official MSA deliniations cut through continuous metro areas because of the presence of a state border. In this case, we ignore the official metro deliniation. Second,

[1] Contact: dwalls2@earthlink.net

Table A.1.1: Aggregation to five industry types

Aggregation	NAISC Code	NAISC Industry name
Manufacturing and wholesale	31-33	Manufacturing
	42	Wholesale
	48-49	Transportation and Warehousing
Non-tradable services, including retail	44-45	Retail
	71	Arts, Entertainment, and Recreation
	72	Accommodation and Food Services
Public services	61	Educational Services
	62	Health Care and Social Assistance
	92	Public administration
Tradable services	51	Information
	52	Finance and Insurance
	53	Real estate
	54	Professional, Scientific, and Technical Services
	55	Management of Companies and Enterprises
Others	11	Agriculture, Forestry, Fishing and Hunting
	21	Mining
	22	Utilities
	23	Construction
	56	Administrative and Support and Waste Management and Remediation Services
	81	Other Services (except Public Administration)

Note: The industries are described in detail at the website of the Census Bureau (`https://www.census.gov/naics/?58967?yearbck=2017`).

Table A.1.2: NETS establishments by industry in selected US cities

MSA	NETS establishments: Mfg and wholesale	Non-tradable services	Public services	Tradable services	Total
Chicago	62,650	75,955	55,128	145,346	521,827
Los Angeles	118,537	159,622	96,941	276,662	991,394
New York	133,398	201,065	124,228	317,759	1,176,488
all other MSAs	2,315,635	3,458,584	2,267,760	6,076,959	24,950,209

Note: Data are from the National Establishment Time Series (NETS) database for the year 2012. See text for details. We do not report the 'residual' category of "Other" establishments.

our criterion sometimes leads to the exclusion of cities/counties that appear in the MSA name. We either move the centroid manually or, where necessary because the Metro is too large for a 70x70 grid (e.g. Dallas-Fort Worth), we enlarge the grid.

A.2 Big data on 125 cities

A.2.1 Sampled cities and city delineation

Sampled cities To get a meaningful metro-area as the unit of analysis, a potential location has to fulfill three conditions: (i) the location must have a minimum number of SNL grade-

A office space buildings such that the algorithm has a chance to identify it; (ii) it should either be of a substantial size itself or belong to a larger metro area of a substantial size; and finally (iii) the locations that are combined as a metro should not only be sufficiently large in terms of the number of buildings but should have some relevance broadly defined (e.g., population, population rank in the system of cities within a country etc). Correspondingly, our selection process works in three stages:

1. We identify metro areas as clusters of nearby "cities" (according to SNL definition) with SNL buildings, to which we merge Starbucks franchises. We define "core cities" as cities (defined by median x and y coordinates) that dominate all other cities within 30 kilometers in terms of the number of establishments (SNL and Starbucks). All non-core cities are then assigned to the nearest core city if a core city is within 30 kilometers. Each cluster of one core and potentially several non-core cities constitutes a metro area (a metro can also be constituted of one core city alone).

2. We keep metros in our sample that belong to a metro area with at least 25 SNL buildings or 25 Starbucks franchises in order to ensure that we have sufficient mass in the metro so that the clusters likely represent meaningful concentrations.

3. We manually process the resulting data set to make sure that the metro name corresponds to conventions, i.e., we choose the name of a well-known non-core city if the core city is less well-known but was selected by the algorithm as a core city because of the larger number of buildings and Starbucks franchises. We drop a small number of identified metro areas which, despite passing the identification thresholds, appear of limited relevance to our analysis (e.g., Princeton, Mountain View, Parsippany).

All 125 cities in our sample conform to these criteria and Figure A.2.1 maps them. Even though we lack coverage for India and large parts of Africa, our sample is truly global in the sense that we have cities of global importance from all continents.

Delineation of city grids To endogenously define the extent of a given city in our sample, we chose the following procedure. We start from the SNL database and drop any office buildings that are more than 30 kilometers away from the city-median (based on the other office locations) in either x or y coordinates. For each city, we then use an algorithm to create midpoints for 250 meters grids for the entire city area spanned by the widest x and y coordinate still found in the SNL data, plus allowing for an additional 5 kilometers buffer around these. This leaves us with 250 meters grids for each city, to which various data sources can be spatially merged.

Overlap with other global city lists Our list of 125 cities shows a high degree of overlap with the list of 123 global cities by Brooking's Global City Initiative (Trujillo and Parilla,

Figure A.2.1: The sample of 125 global cities

Notes: To enter the sample, a city needs to have at least 25 prime investments in grade-A office stock by real estate investment trusts in the SNL-S&P database or at least 25 Starbucks franchises. The full list of cities that we use is provide in the notes to Table 2 and available in the replication archive and our GitHub toolkit.

2016). It also covers 87% of the GDP produced by the cities of the top-100 cities list by PricewaterhouseCoopers (2009).

A.2.2 'Big data' establishments

Data are labeled according to two uses: (i) core dataset; and (ii) validation. The core data are used for the prediction of tradable services employment.

A.2.2.1 Core data for tradable services employment predictions

Google places data We use Google's Nearby Places API to scrape the coordinates of all Places of Interest (POI) associated with specific keywords and located in our sample of global cities (see Table A.2.1 for the associated keywords).

Table A.2.1: Prime Service Establishments by Industry

Type of Firm	Search term(s)	Establishments
Accounting Firm	accountant	58,862
Central Bank	central bank	298
Consultancy	consultant; consultancy	15,251
Insurance	insurance	30,827
Investment Bank	investment bank	2,432
Law Firm	law firm	24,056
Stock Exchange	stock exchange	179
Total		131,905

While Google's API identifies all POI within a circle with a user-specified centroid and radius (with a maximum of 50 kilometers), it returns at most 60 POI per search query. To collect the universe of POI despite this query restriction, we therefore apply an iterative search strategy for each city-keyword pair:

1. We perform an initial scrape on each city's centroid, using a radius of 50 kilometers. If the query returns less than 60 POI, we stop.
2. If the query returns 60 POI, we perform additional scrapes within 4 circles with 25km radius each, shifting their centroids by 25km in each inter-cardinal direction from the original circle's centroid.
3. We continue to divide these circles into 4 overlapping subcircles in the same manner until the respective query returns less than 60 POI.
4. Finally, we delete all duplicates.

As each step of our iterative search strategy generates sub-circles that cover the whole area of their parent, we are guaranteed to obtain the universe of all POI within our area of interest matching the specified keywords. Overall, we end up with a sample of 131,905 establishments. We merge these to our grid dataset, which results in dropping a substantial amount of them that do not lie in our MSA grids.

Office locations of global tradable services companies/establishments Table A.2.2 summarizes our data on establishments of cleanly identified global industry leaders as well as stock markets and central banks. In total, we collected by hand more than 3,200 individual establishment locations for our 125 city sample.

Table A.2.2: Global Prime Service Establishments by Industry

Type	Establishments	Share
Accounting Firms	799	24.6
Consultancies	404	12.4
Investment Banks	467	14.4
Law Firms	360	11.1
Insurance	310	9.6
HQs of listed firms	562	17.3
Stock Exchange	217	6.7
Central Bank	127	3.9
Total	3,246	100.0

As a first type of data, we gather the global office locations of leaders in certain tradable services industries. In particular, we identify the most important global firms in accounting, consultancy, law, investment banking, and insurance based on the *Financial Times* and respective industry magazines. Table A.2.3 names the firms that we consider industry leaders as well as their numbers of establishments in our sample. We retrieve the coordinates of all the offices of the respective companies via their websites. Where coordinates are not available, we retrieve addresses instead. We transform these into coordinates via a self-written Matlab program that queries the Google API (as documented in Table A.2.3).

As a second group of globally important businesses, we collect data on the largest listed tradable services companies in each country that is present in our city dataset. The simple criterion is that the company is part of the country's main stock market index. For example, the insurer Munich RE is a tradable services company, is listed in the DAX 40 (the German leading stock market index), and has its HQ in Munich. In total, we geocode the location of 562 listed tradable services companies.[2]

As a third group of global tradable services establishments, we collect the addresses of all central banks and stock markets of the countries in our sample as well as their domestic regional and international representations.[3]

A.2.2.2 Validation datasets

Coworking spaces We collect the locations of the two leading office space providers, Regus and WeWork, from their websites. We identify 2,253 locations for Regus and 742 for WeWork.

Real estate investments in grade-A office stock These (proprietary) data are provided by SNL-S&P real estate research (see http://www.snl.com/Sectors/RealEstate/).

Starbucks franchises The Starbucks data can be downloaded from the following url (https://opendata.socrata.com/Business/starbucks/cxf4-mc6k). This open source dataset provides the location of all Starbucks shops as scraped from the shop finder from the Starbucks website. We manually checked and corrected data points where necessary.

Tall buildings data The *Emporis* dataset contains information of tall buildings across the world and is typically considered the most comprehensive of its kind. The dataset is on the level of the individual building and includes buildings that no longer exist, which prevents a mechanical survival bias. It contains the geographic coordinates, the construction year, and various building attributes of which the height is the one that is most comprehensively

[2] Only very few of these 562 companies do not have their HQ in one of the cities in our sample (e.g., SAP in Germany).

[3] To identify all potential stock exchanges, we rely on a list compiled by Meri Paterson (http://www.meripaterson.com). We then visit the websites of the exchanges and retrieve their addresses. For the central banks, we visit the websites of the central banks of the countries in our sample. Furthermore, we add the headquarters and international representations of the ECB and the BIS.

Table A.2.3: Global Prime Service Companies and Establishments

Establishment Type & Company name	Source	*N*	Coordinates	Address
		Accounting firms		
Ernest & Young	Company's website	197		X
Deloitte	Company's website	215		X
PWC	Company's website	194		X
KPMG	Company's website	193	X	
		Consultancies		
Accenture	Company's website	118		X
Boston Consulting Group	Company's website	89	X	
McKinsey	Company's website	99	X	
Booz Allen	Company's website	51		X
Bain	Company's website	47	X	
		Law firms		
Kirkland	Company's website	15	X	
Latham	Company's website	25		X
BakerMcKenzie	Company's website	54	X	
DLA Piper	Company's website	65	X	
Skadden	Company's website	22	X	
Dentons	Company's website	71		X
Clifford Chance	Company's website	21	X	
Sidley Austin	Company's website	19	X	
Hogan Lovells	Company's website	39		X
Allen Overy	Company's website	29	X	
		Investment banks		
JP Morgan	Company's website	56	X	
Goldman Sachs	Company's website	59	X	
Merill Lynch	Company's website	49	X	
Morgan Stanley	Company's website	45	X	
Citibank	Company's website	43		X
Barclays	Company's website	31	X	
Credit Suisse	Company's website	57		X
Deutsche Bank	Company's website	74	X	
Wells Fargo	Company's website	24	(X)	
HSBC	Company's website	29		X
		Global insurances		
Allianz	Company's website	128	X	
Axa	Company's website	1377	X	
Prudential Financial	Company's website	45	X	

covered. Another such variable with good coverage is building use, which provides information of the type of building, i.e., whether it contains offices, apartments, hotels, some other functions (e.g. churches, stadiums, etc). This information can be grouped to meta-categories such as residential, commercial, and other (Ahlfeldt and McMillen, 2018).

Social media data Geo-tagged photos come from Eric Fisher's Geotaggers World atlas. Data scientist Eric Fisher obtained a worldwide data set of photos from Picasa and Flickr via the APIs of the respective platforms. We are grateful to him for sharing the data with us. He argues high photo densities reflect places of human interest.

Twitter For about half of the 125 cities in our sample (others were not available), we bought Twitter (now X) data from the social-media company *gnip*, which is now owned by X. The data cover those tweets which were categorized by the provider as relating to business.

B Additional computations, results, and documentation

B.1 Clustering algorithm

Clustering algorithm Using pseudo-code, the figure below describes our algorithm to detect and delineate prime locations (Algorithm 1). It identifies cells that contain abnormally high densities of employment within the city. It then aggregates these 'cluster grid cells' into "orthogonally convex" spatial units that we call prime locations.

Algorithm 1: Prime location identification

Data: City grid; set of plants p_i in the city, with employment and latitude-longitude; undevelopable cells mask

- **begin**
  - **for** *each city* **do**
    - **for** $i \leq 100K$ **do**
      - draw a developable random location l_i within the city
      - compute employment e_i within a disk of radius 750 metres around l_i
      - adjust e_i for the share of undevelopable cells in the disk
    - Compute the 99.5th percentile of the distribution of the 100K e_i
    - **for** *all plants p_i in our city* **do**
      - compute employment $\widetilde{e}_i$ within a disk or radius 750 metres around p_i
      - **if** *$\widetilde{e}_i$ above the 99.5th percentile of e_i distribution* **then**
        - mark p_i as cluster point
    - map all cluster points to grid cells; grid cells with at least one cluster point are called 'cluster grid cells'
    - mark all cluster grid cells as unprocessed, sort them in decreasing order of cluster point weights
    - **while** *there are unprocessed cluster grid cells* **do**
      - pick the unprocessed cluster grid cells with the largest weight, give it a cluster index
      - find all connected cells, give them the same cluster index, mark them as processed
    - rank clusters c_j by decreasing employment
    - **for** *all clusters c_j in the city* **do**
      - **if** *employment in c_j less than $\mathcal{T} = 5\%$ of largest cluster c_1 in city* **then**
        - drop cluster
      - **else**
        - Assign prime location status
    - **while** *Prime location identifier changes* **do**
      - generate pairs of nearest prime locations
      - **if** *centroid-distance $\mathcal{D} \leq 2.5km$ and border-distance $\mathcal{D} \leq 1km$* **then**
        - merge smaller to larger prime location
      - **else**
        - do nothing
    - **for** *each prime location* **do**
      - Generate orthogonal convex hull
      - **if** *vertical or horizontal extent $\geq$ 5km and there is a thin waist within 2.5km of the geographic centroid* **then**
        - Split at the waist
      - **else**
        - do nothing

Result: Prime locations

The algorithm's density cutoff To identify a percentile cutoff to differentiate abnormally high from ordinary densities, we rely on employment gradients. Specifically, we want to detect areas of high densities that are clearly separable from areas with lower densities. We estimate gradients using the following equation: $y_i = \sum_{b=1}^{28} B_b + \gamma_m + \epsilon_i$, where y_i is the employment density (jobs per square kilometer) in grid cell i, B is an indicator variable that indexes the b's distance bin, and γ_m is a metro-specific constant. For all distances $d \in [-0.6\text{km}, 1\text{km}]$, we employ 200 meters bins. For $d > 1\text{km}$, we employ bins of 1 kilometer. Note that we calculate distances for cells lying within prime locations ($d \leq 0$) as well as those lying outside of them ($d > 0$). Figure B.1.1 plots the coefficient for B_b relative to the outermost bin with a 10% confidence interval. These densities are conditional on metro fixed effects.

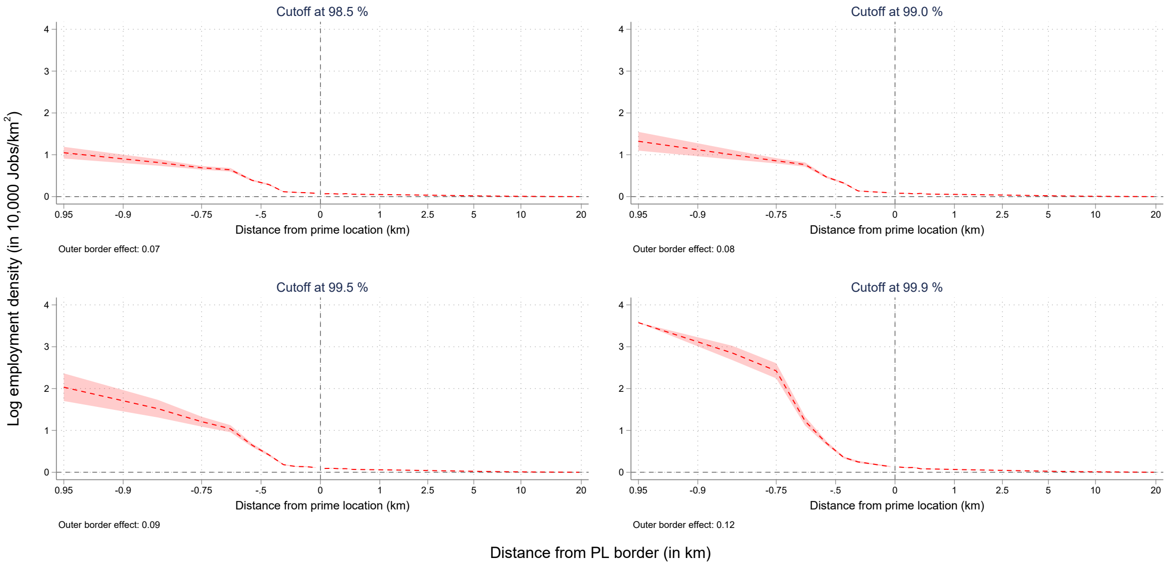
Low cutoffs such as the 98.5th and 99th percentile of the distribution of employment densities across developable cells identify areas that are not dense enough (see the first rows of Figure B.1.1). Too high cutoffs such as the 99.9th percentile result in a 'spatial pre-trend', i.e., density increases stronger in proximity to the prime location border outside the prime location. This is reflected in higher densities (outer border effect) just outside the prime loction (lower right graphs in Figure B.1.1). This effect is particularly pronounced for the 10 largest MSAs which host particularly important prime locations (panel b). We hence settle for a cutoff at the 99.5th percentile, showing both high levels of employment density and a relatively flat spatial pre-trend.

B.2 Prime locations in US metropolitan ares

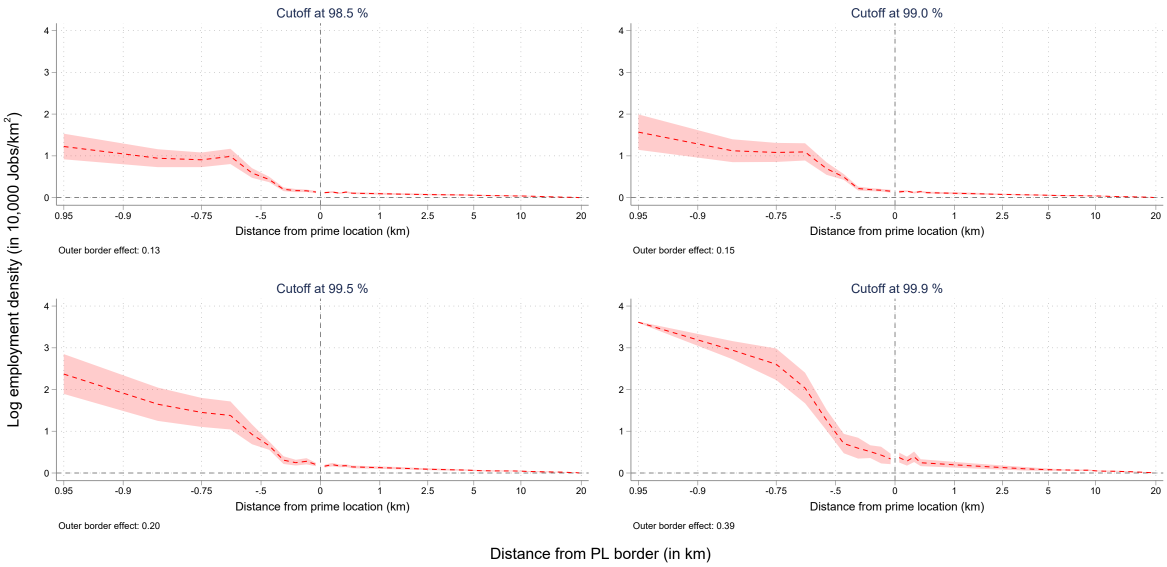
Prime locations and historic manufacturing shares In Table B.2.1, we regress the number of prime locations and their share at total MSA employment against the 1940 share of manufacturing establishments at manufacturing and services employment. We control for city size since the historic manufacturing share is likely correlated with contemporaneous city size. For large MSAs (employment $\geq 1\text{M}$), we find that a greater historic manufacturing share is associated with a significantly smaller number of prime locations and a significantly larger share of prime locations at total employment, suggesting that contemporaneous prime locations are larger in cities that historically specialized in manufacturing.

Prime locations and land-use regulations In Table B.2.2, we regress the number of prime locations and the share at total MSA employment against the 2019 Wharton Residential Land Use Regulatory Index (WRLURI), which we aggregate from the municipality to the MSA level. We control for city size since the restrictiveness of land-use regulation is likely correlated with city size. The key insight is that large cities with more restrictive land-use regulations tend to have fewer prime locations which suggests that zoning might act as a deterrent to the formation of sub-centers. In Table B.2.3, we repeat the analysis using

Figure B.1.1: Employment density gradients at different cutoffs



(a) All US MSAs



(b) Top-10 MSAs

Notes: The gradients are estimated using a binned regression (see text). The densities depicted on the y-axis are conditional on the metro fixed effects and relative to the outermost bin (20-21km).

the adjusted height measure by Barr and Jedwab (2023) as an alternative proxy for land-use regulations. The adjusted height is a residual adjusting the observed sum of heights across all buildings for a battery of variables that capture demand for height. It can thus be interpreted as an inverse measure of height restriction. As expected, cities that build more—especially in the vertical dimension because of less stringent height restrictions—tend to have fewer, but stronger, prime locations.

Table B.2.1: Prime location geography and historic manufacturing share

	(1)	(2)	(3)	(4)	(5)	(6)
	Number of prime locations			PL share at total employment (%)		
MFG estab. share in 1940 (%)	-0.013	-0.009	-0.183	-0.052	-0.048	0.528
	(0.0045)	(0.0037)	(0.0749)	(0.0301)	(0.0276)	(0.1389)
Log employment		0.689	1.990		0.836	2.230
		(0.0913)	(0.7150)		(0.2161)	(1.3027)
MSAs	All	All	Emp. $\geq$ 1M	All	All	Emp. $\geq$ 1M
Observations	367	367	35	367	367	35
R^2	.0104	.463	.349	.00938	.0454	.463

Notes: Unit of observation is MSA. MFG share is the share of manufacturing establishments at manufacturing and services and establishments in an MSA. MSA size is measured in terms of 2015 total employment. Columns (1), (2), (4), and (5) report results for all MSAs, whereas columns (3) and (6) report results for large MSAs with $\geq$ 1M employment (in terms of 2015 total employment). The manufacturing data by county come from the 1940 Census of "Population, Housing, Agriculture & Economic Data," avaible from IPUMS NHGIS, University of Minnesota. Counties are aggregated to MSAs using the crosswalk provided by the Census Bureau and NBER.

Table B.2.2: Prime location geography and land use regulation

	(1)	(2)	(3)	(4)	(5)	(6)
	Number of prime locations			PL share at total employment (%)		
WRLURI	0.236	-0.048	-1.105	0.471	0.143	-0.946
	(0.0789)	(0.0512)	(0.6124)	(0.3953)	(0.4070)	(1.2474)
Log employment		0.743	1.891		0.856	3.002
		(0.1003)	(0.8823)		(0.2385)	(1.6741)
MSAs	All	All	Emp. $\geq$ 1M	All	All	Emp. $\geq$ 1M
Observations	303	303	35	303	303	35
R^2	.0204	.461	.253	.00519	.0428	.146

Notes: Unit of observation is MSA. PL share at total employment is the percent of city-wide employment located in all the MSA's prime locations. Columns (1), (2), (4), and (5) report results for all MSAs, whereas columns (3) and (6) report results for large MSAs with $\geq$ 1M employment (in terms of 2015 total employment). WRLURI denotes the Wharton Residential Land Use Regulatory Index 2019 (Gyourko et al., 2021), aggregated to MSAs.

Table B.2.3: Prime location geography and adjusted building height

	(1)	(2)	(3)	(4)	(5)	(6)
	Number of prime locations			PL share at total employment (%)		
Adjusted height	-0.021	-0.045	-0.429	0.604	0.578	1.370
	(0.0269)	(0.0235)	(0.4122)	(0.1541)	(0.1543)	(0.4911)
Log employment		0.877	1.872		0.907	2.530
		(0.1257)	(0.8319)		(0.2998)	(1.5334)
MSAs	All	All	Emp. $\geq$ 1M	All	All	Emp. $\geq$ 1M
Observations	216	216	35	216	216	35
R^2	.00127	.472	.249	.0932	.136	.26

Notes: Unit of observation is MSA. PL share at total employment is the percent of city-wide employment located in all the MSA's prime locations. Columns (1), (2), (4), and (5) report results for all MSAs, whereas columns (3) and (6) report results for large MSAs with $\geq$ 1M employment (in terms of 2015 total employment). Adjusted height is the sum of the height of all tall buildings in a city controlling for demand factors from Barr and Jedwab (2023), aggregated to MSAs.

Prime location specialization In Table B.2.4 we compute the Herfindahl-Hirschman Index (HHI) of industry sector specialization by prime location and regress it against either the number of prime locations within an MSA or the employment rank of a prime location within an MSA. To account for heterogeneity, we also conduct separate regressions for small and large MSAs. In small MSAs, we observe a higher degree of sectoral specialization of prime locations as the number of prime locations increases. In the regressions against the number of prime locations, we control for the HHI outside prime locations to account for any correlation between spatial structure and industry concentration. The rank effect is estimated from within-MSA variation by adding MSA fixed effects. We find the opposite for large MSAs, which suggests a role for agglomeration economies: the larger a prime location grows, the more it generates agglomeration economies within a specific sector, making it more difficult for other sectors to afford high rents associated with sector-specific productivity. For large cities, we also find that lower-ranking prime locations are significantly more specialized, suggesting that one reason for secondary prime locations to emerge is to capitalize on sector-specific agglomeration economies For small cities, the estimated rent effect is qualitatively the same, but it is not significant at conventional levels.

Table B.2.4: Prime location specialization

	(1)	(2)	(3)	(4)	(5)	(6)
	Herfindahl–Hirschman Index (HHI) [0,1]					
# prime locations (PL)	-0.002		0.086		-0.013	
	(0.0044)		(0.0129)		(0.0048)	
HHI, outside PL	-0.135		0.490		1.091	
	(0.6644)		(0.6941)		(2.5684)	
PL rank		0.035		0.123		0.026
		(0.0185)		(0.0990)		(0.0094)
Constant	0.443	0.352	0.194	0.271	0.232	0.320
	(0.1481)	(0.0301)	(0.1580)	(0.1142)	(0.5462)	(0.0282)
MSA effects	-	Yes	-	Yes	-	Yes
MSAs	All	All	Emp. < 1M	Emp. < 1M	Emp. $\geq$ 1M	Emp. $\geq$ 1M
Observations	531	531	395	395	136	136
R^2	.00032	.686	.064	.802	.0325	.307

Notes: Unit of observation is prime location. Herfindahl–Hirschman Index is an index of sector specialization by prime location calculated using the following sectors: Manufacturing and wholesale, tradable services, non-tradable services, public services, other sectors. Standard errors (in parentheses) clustered on MSAs.

Prime locations and congestion measures In Table B.2.5, we correlate the number of prime locations and their share at total MSA employment with measures of housing costs, accessibility, and commuting time. Since both outcome measures are correlated with city size, we control for the latter. We do not find a significant conditional correlation between the MSA house price index and either of the outcome measures. As expected, the nearest prime location is closer in MSAs with more prime locations. The commuting time, however, is not lower in MSAs with more prime locations. On the contrary, commuting times are significantly lower in MSAs where prime locations are larger, which are those that have fewer prime locations. This suggests a role for scale economies in public transport infrastructure, especially mass transit: it may be relatively more efficient to connect strong prime locations with a hub-and-spoke subway system.

Table B.2.5: Prime locations and congestion measures

	(1)	(2)	(3)	(4)	(5)	(6)
	Ln house price index		Ln PL accessibility		Ln commuting time	
Log employment	0.146	0.127	0.224	0.199	0.050	0.058
	(0.0264)	(0.0206)	(0.0225)	(0.0156)	(0.0085)	(0.0059)
Number of prime locations	-0.021		-0.042		0.009	
	(0.0249)		(0.0167)		(0.0055)	
PL share at total employment (%)		0.005		-0.003		-0.003
		(0.0043)		(0.0033)		(0.0015)
Observations	368	368	381	381	381	381
R^2	.117	.118	.271	.266	.188	.196

Notes: Unit of observation is MSA. House price index is the *Zillow ZHVI Singe-Family Homes Tie Series for Metros*, downloaded on October 15, 2024. PL accessibility is the share of the population living within 20 km of the nearest prime location, computed from census block group data. Commuting time is the population-weighted average commuting time computed from census block group data. City size is measured in terms of employment. Robust standard errors in parentheses.

Prime location gradients In Table B.2.6, we estimate density, income, and housing cost gradients by regressing outcome measures in logs against distance from the nearest prime locations. To avoid overweighting of larger distances, we aggregate census block group data to 1km-distance bins by prime location catchment area (block groups sharing the same nearest prime location). Consistent with the predictions of the monocentric city model, both population density and the density of housing units decrease by 6.4% per kilometer distance from the nearest prime location. The income gradient is positive, especially with respect to the more central dominant prime location, revealing that, on average, US suburbs tend to be rich. This positive income gradient implies that households consume more housing space in the suburbs, which may explain why rent and housing value gradients are close to zero (though statistically significant). For the purpose of estimating the rent gradient, it is an unfortunate limitation that the census only reports prices of units and not prices per unit of floor space.

Prime location characteristics and variation in spatial structure Figures B.2.1 and B.2.2 complement Table B.2.7 by illustrating the distribution of selected prime location characteristics and by summarizing the geography of prime location by US MSAs.

Validation using County Business Pattern data In Figure B.2.3, we subject our prime locations delineated based on NETS data to a validation exercise using official County Business

Table B.2.6: Prime location gradients in US MSAs

	Ln pop. dens.	ln housing dens.	ln income	ln rent	ln housing value
Panel (a): All prime locations					
Distance from PL (km)	-0.064	-0.064	0.001	-0.005	0.008
	(0.0013)	(0.0014)	(0.0003)	(0.0003)	(0.0004)
Observations	21851	21842	21827	20942	21811
R^2	.571	.56	.37	.646	.196
Panel (b): Prime locations with MSAs with employment $\geq$ 1M (large MSAs)					
Distance from PL (km)	-0.070	-0.072	0.005	-0.002	0.015
	(0.0026)	(0.0027)	(0.0010)	(0.0007)	(0.0009)
Observations	5260	5259	5255	5076	5253
R^2	.657	.665	.384	.592	.341
Panel (c): Prime location rank within MSA = 1 (large MSAs)					
Distance from PL (km)	-0.074	-0.079	0.013	0.002	0.021
	(0.0053)	(0.0056)	(0.0012)	(0.0011)	(0.0013)
Observations	1540	1539	1538	1491	1538
R^2	.703	.705	.47	.618	.442
Panel (d): Prime location rank within MSA $\geq$ 2 (large MSAs)					
Distance from PL (km)	-0.068	-0.070	0.002	-0.004	0.012
	(0.0029)	(0.0031)	(0.0011)	(0.0008)	(0.0010)
Observations	3720	3720	3717	3585	3715
R^2	.636	.649	.385	.594	.312

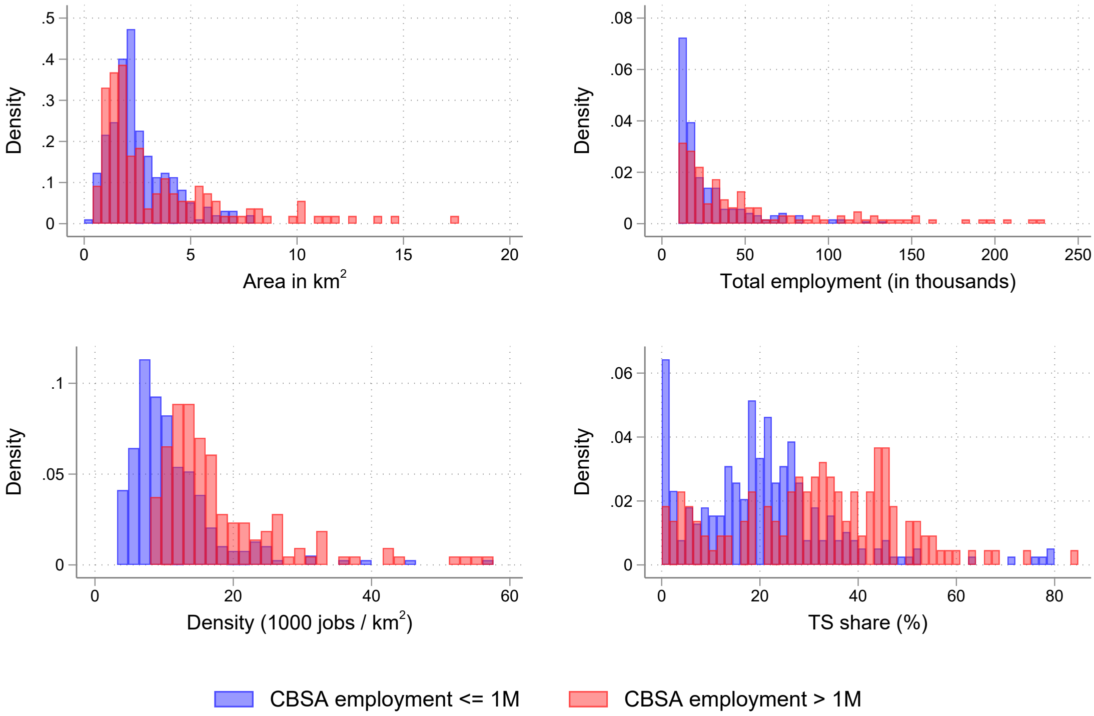
Notes: Unit of observation is 1km-distance from prime location bin by prime location catchment area (observations sharing the same nearest prime location). Bins are aggregated from census block groups and restricted to bins up to 50 km from the nearest prime location. All models include prime location catchment area effects. Standard errors (in parentheses) are clustered on prime location catchment areas. The blockgroup data for population, housing units, per capita income, rents, and housing values at the blockgroup level are from the 2010 Census, avaible from IPUMS NHGIS, University of Minnesota. The county surface areas have been computed from the Census Tiger shapefiles.

Table B.2.7: Summary of spatial structure

			Prime locations' share (%) at		
CBSA	Classification	# PL	area	employment	TS emp.
New York	Duocentric-Agglomerated	2	0.44%	25.43%	48.09%
Los Angeles	Polycentric-Dispersed	9	0.53%	9.42%	17.69%
Miami	Polycentric-Dispersed	8	0.19%	7.66%	11.66%
Chicago	Monocentric-Agglomerated	1	0.36%	17.48%	37.54%
Dallas	Polycentric-Agglomerated	9	0.17%	13.06%	25.41%
Washington	Polycentric-Dispersed	4	0.37%	20.27%	21.51%
Houston	Polycentric-Dispersed	8	0.14%	13.87%	21.26%
Philadelphia	Polycentric-Agglomerated	3	0.38%	16.78%	24.65%
Atlanta	Polycentric-Agglomerated	4	0.49%	14.11%	23.09%
Boston	Monocentric-Agglomerated	1	0.46%	20.23%	35.85%
Mean, employment $\geq$ 1M		3.78	0.28%	13.27%	21.22%
Mean, employment $<$ 1M		1.158	0.10%	10.76%	14.85%

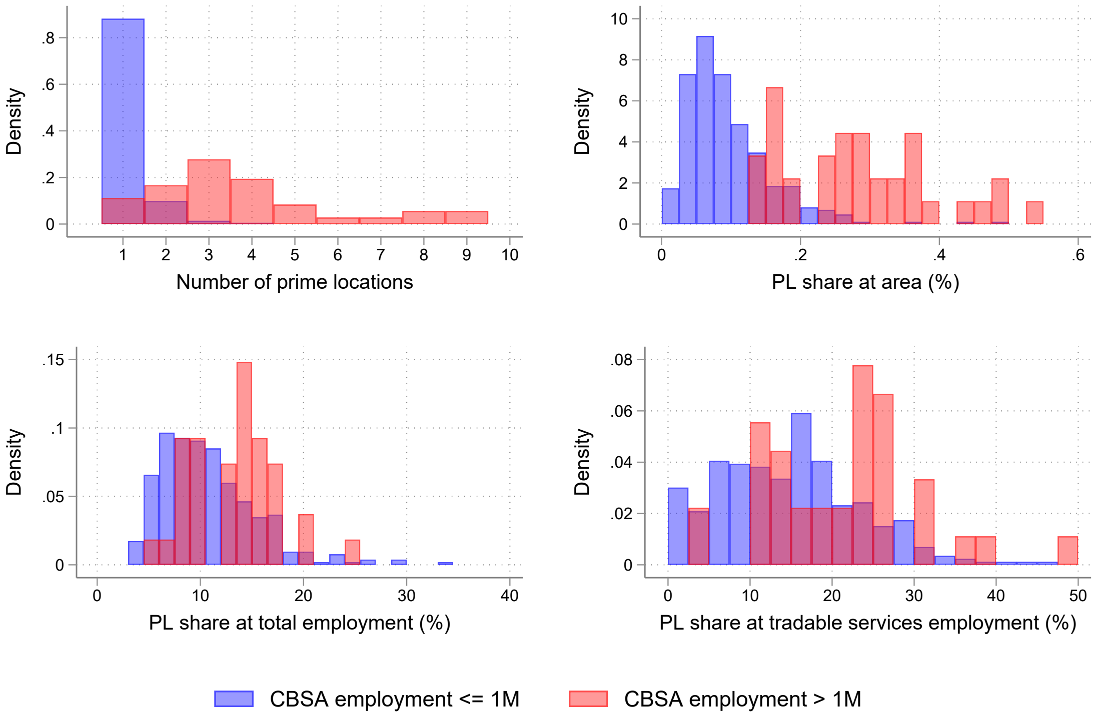
Notes: Agglomerated and dispersed cities are defined by on whether the share of prime locations at total employment is larger or smaller than the median share. We define an MSA as agglomerated if prime locations' share at MSA's tradable services employment is above the median within its size category (employment $\geq$ 1M or < 1M). A full-length version of this table for 381 MSAs is provided in our GitHub toolkit

Figure B.2.1: Characteristics of prime locations



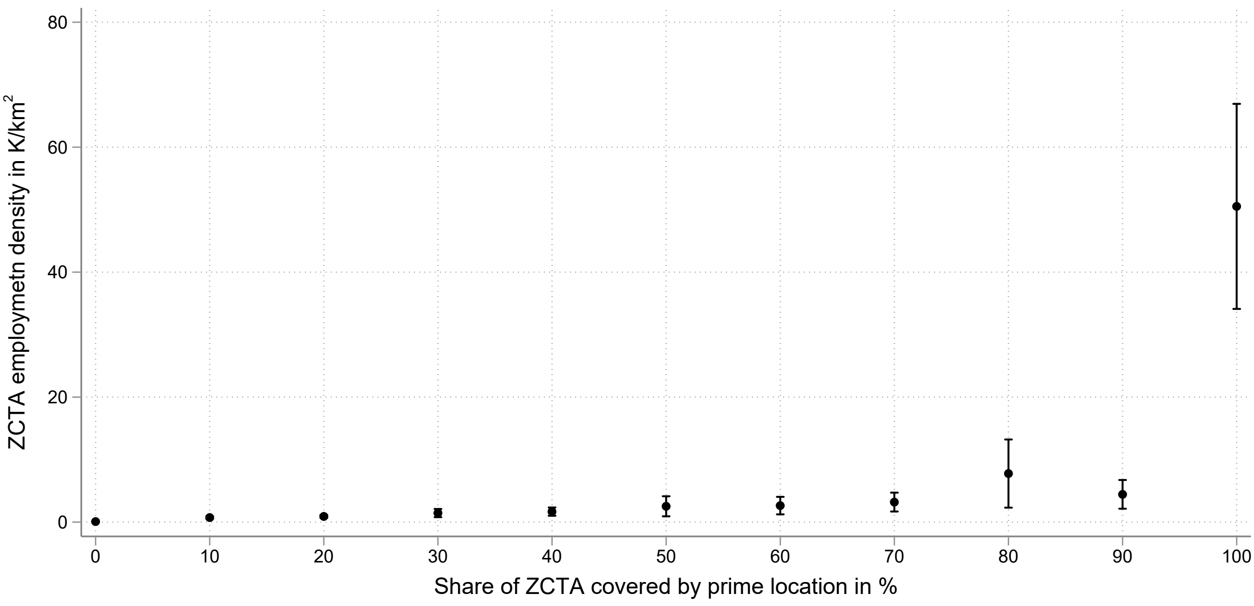
Notes: Unit of observation is prime location. To improve readability, we drop outliers that exceed the following values: Area $\geq$ 50 km^2, employment $\geq$ 250k, employment density $\geq$ 100k.

Figure B.2.2: Importance of prime locations



Notes: Unit of observation is MSA.

Figure B.2.3: Employment density within prime locations in County Business Pattern



Notes: Confidence bands are at the 95% level. We aggregate shares (on the x-axis) to bins defined as follows: 0-5%, 5%-15%, ..., 95%-100%.

Pattern (CBP) registry data. To this end, we intersect official zip code tabulation areas (ZCTAs) and our delineated prime location boundaries within a Geographic Information System (GIS) and merge the CBP zip code employment to the ZCTAs using an offical crosswalk. We then compute the means and the standard errors of the distributions of employment density across ZCTAs within 10-percentage point bins defined in terms of shares of ZCTAs' geographic area covered by prime locations. Employment densities are very low in ZCTAs that do not intersect with prime locations. Employment densities are generally higher when a greater share of ZCTA geographic area falls inside prime locations For ZCTAs that are almost fully within prime locations (share $\geq$ 95%), we find extremely high employment densities of about 50K jobs per square kilometer in the CBP data. Hence, based on official CBP data, we can confirm that our delineation algorithm picks locations with ultra-high employment densities, as intended.

B.3 Prime locations within 125 cities

B.3.1 Estimated employment weights

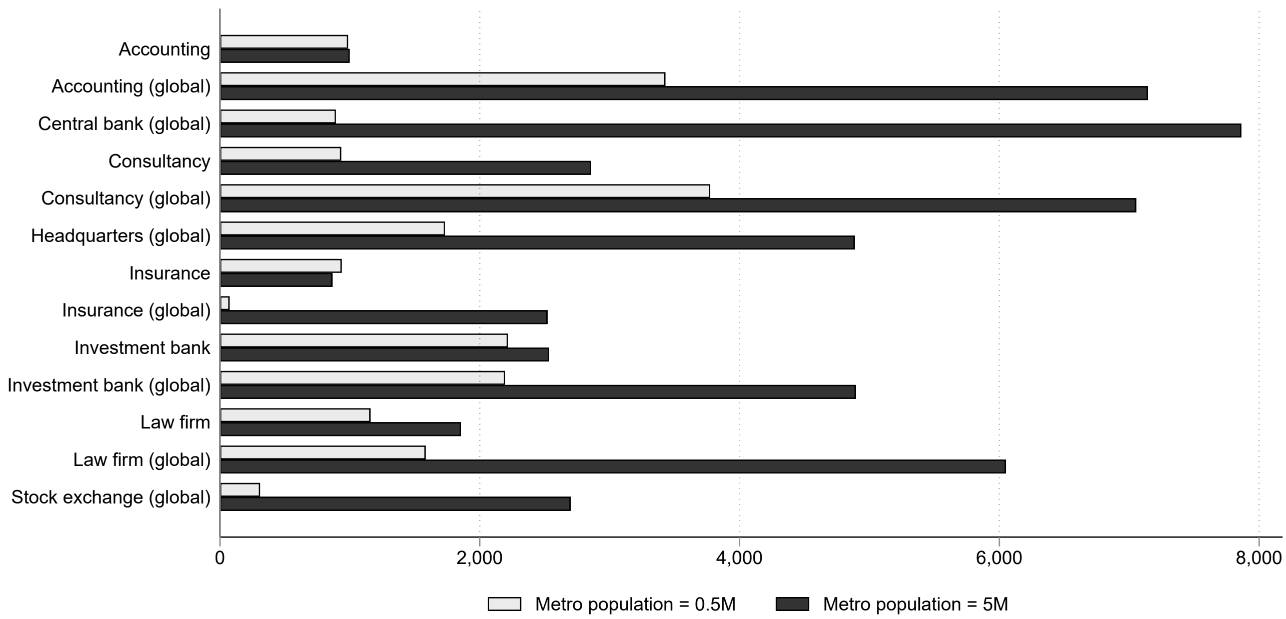
We present the employment weights we assign to establishments queried from the Google Places API and scraped from the web in Table B.3.1 and Figure B.3.1.

Table B.3.1: Employment weights: All MSAs

Establishment type	Constant	S.E.	City size elasticity	S.E.	Pseudo R^2
Accounting	6.827	0.103	0.005	0.007	0.000
Consultancy	0.466	0.066	0.486	0.004	0.378
Insurance	7.305	0.110	−0.035	0.007	0.001
Investment bank	6.943	0.066	0.058	0.004	0.004
Law firm	4.376	0.079	0.204	0.005	0.031
Accounting (global)	3.958	0.041	0.319	0.003	0.065
Central bank (global)	−5.605	0.040	0.945	0.003	0.232
Consultancy (global)	4.672	0.041	0.272	0.003	0.034
Insurance (global)	−15.794	0.071	1.532	0.005	0.370
Investment bank (global)	3.124	0.050	0.348	0.003	0.025
Law firm (global)	−0.271	0.046	0.582	0.003	0.096
Stock exchange (global)	−6.608	0.069	0.941	0.004	0.133
Headquarters (global)	1.548	0.050	0.450	0.003	0.091

Notes: In the first step we estimate employment weights by MSA and establishment type by regressing employment against establishment counts within randomly drawn disks with 750m radius using non-negative least squares method following Lawson and Hanson (1974). In the second step we estimate the city size elasticity by regressing employment weights against ln CBSA population using PPML.

Figure B.3.1: Estimated employment weights by city size: All MSAs



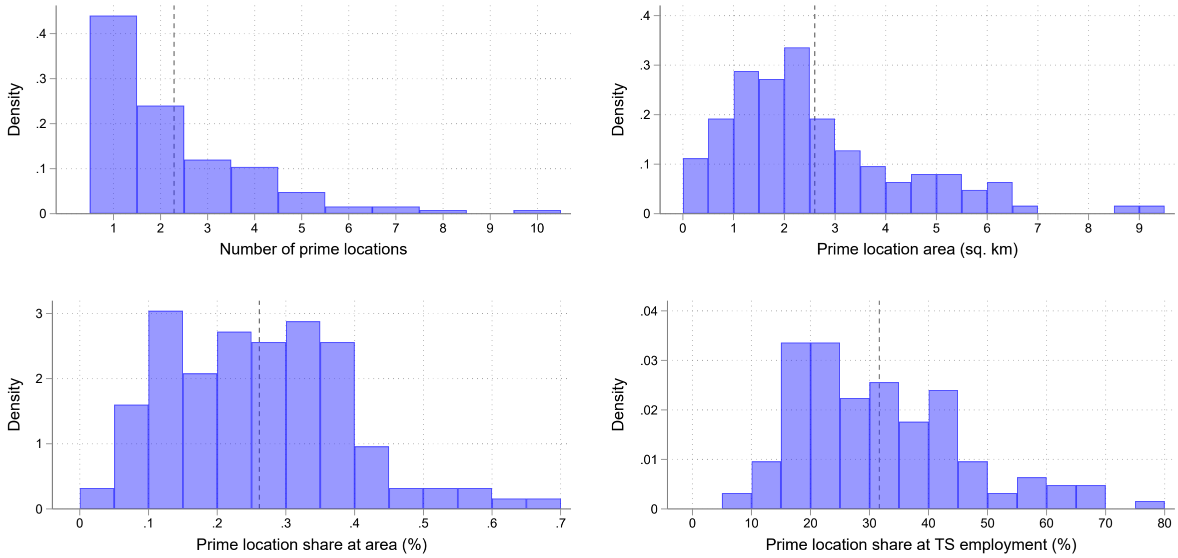
B.3.2 Summary statistics

Figure B.3.2 illustrates the distributions of key features of the 286 prime locations we delineate in the Global Cities sample.

B.3.3 Validation

Delineating prime locations with alternative employment measures Table B.3.2 presents the results of a grid-level regression aimed at measuring the overlap between the delineation

Figure B.3.2: Prime locations in our global sample of 125 cities



Notes: Each vertical line shows the mean of a distribution. Area is the developable area, excluding water and steep slopes. The underlying employment data are our prediction based on establishments from the Google Places API and Web Scraping as well as the employment weights reported in Table B.3.1. In total, we identify 286 prime locations in 125 cities.

of prime locations based on two types of data. The dependent variable is a dummy variable indicating our baseline prime locations, as defined by total employment, which is discussed in Section 2.2. In column (1), we use a similar dummy variable created with the same delineation algorithm, but we restrict the sample to establishments in tradable services. Column (2) further narrows the sample to establishments in sectors where the NAICS industry classification corresponds to our 'big data' search terms. Across all grid cells in the 381 MSAs, the baseline probability of being a prime location is virtually zero. The probability of a grid cell being in one of our baseline prime locations, conditional on being in a prime location delineated by alternative employment measures, is about 75%. This confirms that our search terms, which capture tradable services, are effective in identifying establishments that cluster in prime locations.

Big data establishments vs. micro-geographic employment data In columns (3) and (4) of Table B.3.2, we subject our entire process of collecting 'big data' and assigning employment weights to a rigorous overidentification test. We focus on the 39 MSAs for which we have collected NETs and big data establishments. The sample of MSAs is split into two halves alphabetically. In column (3), we delineate prime locations using employment weights estimated from the first half, which are then used to predict employment in the second half. In column (4), we estimate weights using the second half and use them to predict employment in the first half. In both cases, the predicted employment is input into our delineation algorithm, with the same parameters as in the baseline. Conditional on being in a prime location delineated based on predicted employment, the probability of also being in a prime location delineated using actual employment is 71.2 % and 82.4% in the two samples. Given the large number of grid cells and small size of prime locations, these probabilities are remarkably high, considering that the training and testing data sets are mutually exclusive.

Table B.3.2: Validation of prime location detection

	(1)	(2)	(3)	(4)
	Prime location, baseline measure using all employment			
	Tradable services	Search terms	EWPPs, first batch	EWPPs, second batch
Prime location, proxy	0.760	0.746	0.712	0.824
	(0.0038)	(0.0035)	(0.0141)	(0.0130)
Constant	0.001	0.000	0.002	0.002
	(0.0000)	(0.0000)	(0.0000)	(0.0000)
Observations	18,352,344	18,352,344	1,325,644	1,970,981
R^2	.355	.42	.15	.133

Notes: Unit of observation is the grid cell. Undevelopable grid cells are excluded. Prime location, baseline measure is a dummy taking the value of one if a grid cell belongs to at prime location, and zero otherwise, where the delineation is based on all employment as discussed in Section 2.2. Prime location, proxy measure is a similar dummy variable based on the same delineation algoritihm, but using a proxy measure for total employment.

Validation across 125 cities using proxies for dense employment clusters Table B.3.3 provides a validation for the 125 city sample using gridded point-pattern data on coworking spaces, grade-A office stocks, and Starbucks franchises.

Table B.3.3: Validation of big data establishments

Outcome	Stat.	Africa[a]	Asia	Australia	Europe	N.A.[b]	S.A.[c]
Coworking spaces	Coeff.	0.016	0.048	0.029	0.032	0.036	0.026
	S.E.	0.002	0.007	0.002	0.002	0.003	0.003
	R2	0.247	0.275	0.797	0.487	0.323	0.315
SNL-S&P investments	Coeff.	0.111	0.134	0.069	0.068	0.053	0.037
	S.E.	0.019	0.026	0.010	0.005	0.006	0.005
	R2	0.400	0.287	0.613	0.262	0.101	0.572
Starbucks	Coeff.	0.012	0.069	0.008	0.027	0.074	0.009
	S.E.	0.002	0.009	0.000	0.001	0.005	0.003
	R2	0.060	0.130	0.617	0.375	0.266	0.045
# cities		6	22	3	34	55	5

Notes: [a]Including Middle East. [b]North America. [c]South America. The table reports regression coefficients and within-R^2 from regressions of the count of a given outcome against the count of big data establishments conditional on city fixed effects. Outcomes and big data establishments are measured as counts within 100K randomly drawn disks of 750 meters radius in each city. We discard disks with zero counts in both outcomes *and* big data establishments. In each column, we pool all global cities within a continent in one regression. Standard errors are clustered at the level of 0.01×0.01 latitude-longitude grid cells.

B.4 Variation in urban structure

Prime locations and tradable services employment concentration by continent Table B.4.1 summarizes the number of prime locations per city as well as their shares at city area and tradable-services employment by continent. A full table for all cities is available in our Global Cities appendix.

Table B.4.1: Number and importance of prime locations by world region

Region	Cities	PLs per city	PLs' share at area	PLs' share at TS employment
Africa & Middle East	6	2.00	0.33%	29.06%
Asia	22	2.00	0.25%	37.53%
Australia	3	1.33	0.34%	36.61%
Europe	34	1.35	0.34%	31.34%
North America	55	3.15	0.20%	27.89%
South America	5	1.40	0.33%	49.26%

Notes: Prime location (PL) share at tradable services (TS) employment is the share of prime locations at total predicted tradable services employment.

Prime location gradients by city structure Figure B.4.1 replicates Figure 3 distinguishing between cities by their type of spatial structure (mono-, duo-, or polycentric). Across outcomes, it is apparent that the role of prime locations as nuclei of business activity gradients is independent of whether a particular prime location is located in a mono-, duo-, or polycentric city. Conditional on the city-specific constant, we do not detect significant differences in the gradients across city types, except for grade-A office stock investment density, coworking spaces, and social media activity, where gradients are somewhat flatter and less discontinuous in polycentric cities (yet, even in those cases, the discontinuity is still pronounced). Generally, the steeper gradients and larger discontinuities at the prime locations' borders for mono- and duo-centric cities—regarding outcomes related to business activity—echo our findings on sub-additivity.

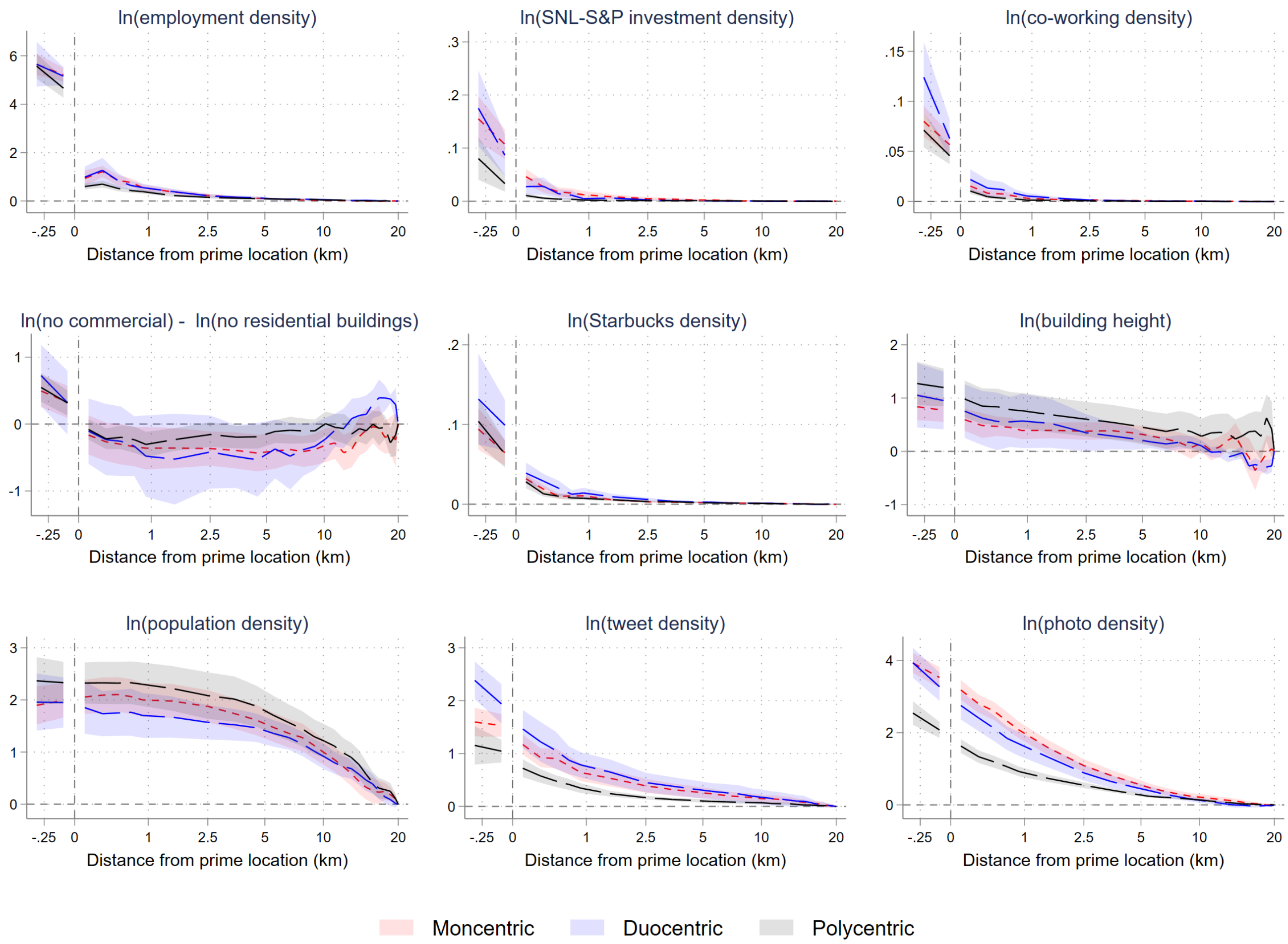
Appendix References

Ahlfeldt, Gabriel M. and Daniel P. McMillen, "Tall buildings and land values: Height and construction cost elasticities in Chicago, 1870-2010," *The Review of Economics and Statistics*, 2018, *100* (5), 861–875.

PricewaterhouseCoopers, " UK Economic Outlook (November 2009) – Which are the largest city economies in the world and how might this change by 2025?," 2009.

Trujillo, Jesus Leal and Joseph Parilla, *Redefining global cities*, The Brookings Institution, 2016.

Figure B.4.1: Prime location distance gradients by spatial structure



Notes: To calculate gradients, we run the regression $y_i = \sum_{b=1}^{27} B_b + \gamma_m + \epsilon_i$, where y_i is outcome y in grid cell i, B is an indicator variable that indexes the b's distance bin, and γ_m is a metro-specific constant. For all distances $d \in [-0.4\text{km}, 1\text{km}]$, we employ 200 meters bins. For $d > 1$km, we employ bins of 1 kilometer. Note that we calculate distances for cells lying within prime locations ($d \leq 0$) as well as those lying outside of them ($d > 0$). The dashed line marks the border of the prime location. Negative distances on the x-axis (left of the dashed line) describe the distance to the border of points within the prime location, positive distances (right of the dashed line) describe the distance to the border of points lying outside of the prime location. Shaded areas represent 90% confidence intervals of the respective point estimates. The gradients are estimated for up to 267 prime locations in 125 global cities, depending on data availability for the particular outcome. The underlying employment data is our prediction based on our 'big data' establishment dataset. For a description of sources for the other outcomes, see Appendix A.2.2.2.