

Online Appendix for “Prime locations”

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

Conditionally accepted, *American Economic Review: Insights*

A Data

We provide detailed information on all the data sources in Table A.2.4 for the US MSAs and in Table A.2.5 for the global cities.

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 70×70 km grid (with 280×280 cells of 250×250 meters) around this centroid. We intersect the resulting grid with the official MSA delineation and only keep the intersecting cells. We then use land cover and digital elevation model (DEM) data to flag undevelopable cells as those that are a majority of water or that

¹Contact: dwalls2@earthlink.net

Table A.1.1: Aggregation to five industry types

Aggregation	Code	NAISC Industry name
Manufacturing and wholesale	31-33 42 48-49	Manufacturing Wholesale Transportation and Warehousing
Non-tradable services, including retail	44-45 71 72	Retail Arts, Entertainment, and Recreation Accommodation and Food Services
Public services	61 62 92	Educational Services Health Care and Social Assistance Public administration
Tradable services	51 52 53 54 55	Information Finance and Insurance Real estate Professional, Scientific, and Technical Services Management of Companies and Enterprises
Others	11 21 22 23 56 81	Agriculture, Forestry, Fishing and Hunting Mining Utilities Construction Administrative and Support and Waste Management and Remediation Services 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				Total
	Mfg and wholesale	Non-tradable services	Public services	Tradable services	
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	2,324,304	3,475,272	2,277,908	6,087,907	25,017,180

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.

have steep slope. For a small number of MSAs, we deviate from this procedure slightly. First, in a very few cases, official MSA delininations cut through continuous metro areas because of the presence of a state border. In this case, we ignore the official metro deliniation. Second, 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

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](#).

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 Brookings’s Global City Initiative ([Trujillo and Parilla, 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).

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:

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

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.

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).

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 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.sn1.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.

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

³To 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	<i>N</i>	Coordinates	Address
<i>Accounting firms</i>				
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	
<i>Consultancies</i>				
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	
<i>Law firms</i>				
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	
<i>Investment banks</i>				
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
<i>Global insurances</i>				
Allianz	Company's website	128	X	
Axa	Company's website	137	X	
Prudential Financial	Company's website	45	X	

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 covered. Another such variable with good coverage is building use, which provides informa-

tion 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, 2018a](#)).

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.

Table A.2.4: Source Overview US MSA Data

Use (/ variable)	Dataset (name)	Provider/originator incl. dataset citation	Data DOI / URL (last accessed)	Related publication	raw (1)	Format in replication package grid (2)	reason for (2)
<i>Data underlying the classification of prime locations</i>							
Generation of establishment-level employment	National Establishment Time-Series (NETS)	Originator: Walls & Associates; transformed version provided by Ahlfeldt et al. (2025b)	proprietary data (Aug 15, 2024)	Walls (2007)	×	×	proprietary data
Definition of MSA grids extends	Prime Locations Toolkit	Ahlfeldt et al. (2025a)	URL (Feb 25, 2025)	Ahlfeldt et al. (2020)	×	×	grid is required format
Landcover (for land developability classification)	ESA WorldCover 10 m 2020 v100	European Space Agency (2022)	DOI (Feb 25, 2025)	Zanaga (2022)	×	×	file size
Slope (for land developability classification)	Global Multi-resolution Terrain Elevation Data 2010	Earth Resources Observation And Science Center (2017b)	DOI (Feb 25, 2025)	Danielson and Gesch (2011)	×	×	manual GIS processing
<i>Correlates and validation</i>							
Adjusted height measure	Adjusted height dataset	Originator: Barr and Jedwab (2023) ; distributor: Ahlfeldt et al. (2025b) (directly shared by authors)	DOI (Feb 27, 2025)	Barr and Jedwab (2023)	×		
CBSA house price index	Zillow ZHVI Single-Family Homes Time Series	Zillow Research (2024a)	URL (Oct 15, 2024)	Zillow Research (2024b)	×		
Land use regulation index	Wharton Residential Land Use Regulatory Index (WRLURI)	Originator: Gyourko et al. (2021) ; distributor: Wharton School (2020)	URL (Feb 18, 2025)	Gyourko et al. (2021)	×		
Historical manufacturing data for 1940	1940 Census - Population, Housing, Agriculture & Economic Data; NHGIS code: 1940_cPHAE	National Historical Geographic Information System (2024)	DOI (Feb 25, 2025)	Steven Manson and Ruggles (2024)	×	×	
	2013 American Community Survey: 5-Year Data; NHGIS code: 2009_2013_ACS5a	National Historical Geographic Information System (2024)	DOI (Feb 25, 2025)	Steven Manson and Ruggles (2024)	×	×	
Generation of gradients for population density, housing density, income, rent, housing values	2010 American Community Survey; NHGIS code: 2010_ACS1	National Historical Geographic Information System (2024)	DOI (Feb 25, 2025)	Steven Manson and Ruggles (2024)	×	×	
Commuting times	County Business Patterns: 2014	United States Census Bureau (2016a)	URL (Feb 25, 2025)	United States Census Bureau (2016b)	×	×	

Notes: *We use the crosswalks from ICPSR Codes to FIPS (Behnam, 2020) and for FIPS to CBSA (Census Bureau, 2023) to spatially match the data.

[†]We use the ZIP-ZCTA crosswalk initially created by John Snow, Inc and made available on github by [Chen and Wasserman \(2024\)](#).

Table A.2.5: Source Overview Global Cities

Use (/variable)	Dataset (name)	Provider/originator incl. dataset citation	Data DOI / URL (last accessed)	Related publication	Format in replication package raw (1) grid (2)
<i>Data underlying the classification of prime locations</i>					
Location of local tradable service establishments	Tradable service firms according to google places	Ahlfeldt et al. (2025b)	DOI (Feb 27, 2025)*	this paper	×
Location of global tradable service firms	Location of industry leaders, headquarters, central banks, stock exchanges	Ahlfeldt et al. (2025b)	DOI (Feb 27, 2025)*	this paper	×
Definition of city grids extends	Prime Locations Toolkit	Ahlfeldt et al. (2025a)	URL (Feb 25, 2025)	Ahlfeldt et al. (2020)	× grid is required format
Landcover (for land developability classification)	Global land cover classification	Earth Resources Observation And Science Center (2017a)	DOI (Feb 25, 2025)	Hansen et al. (2000)	× manual GIS processing
Water bodies (for land developability classification)	Global, high-resolution (30-m) inland water body dataset	United States Geological Survey (2016)	URL (Feb 25, 2025)	Feng et al. (2016)	× manual GIS processing
Slope (for land developability classification)	Global Multi-resolution Terrain Elevation Data 2010 (GMED2010)	Earth Resources Observation And Science Center (2017b)	DOI (Feb 25, 2025)	Danielson and Gesch (2011)	× manual GIS processing
<i>Correlates and validation</i>					
Tweet density	Business-related tweets	Originator: GNIP (now part of X); transformed version provided by Ahlfeldt et al. (2025b)	DOI (Feb 27, 2025)	-	× proprietary data
Geotagged photos	The Geotaggers' World Atlas	Originator: Fischer (2016); distributor: Ahlfeldt et al. (2025a)	URL (Feb 25, 2025)	Fischer (2016)	× manual GIS processing
Emporis building heights data	Emporis building heights data	Originator: Emporis; distributor: Ahlfeldt and McMillen (2017)	DOI (Feb 25, 2025)	Ahlfeldt and McMillen (2018b)	× manual GIS processing
Starbucks establishments	All Starbucks Locations In The World	Guette (2017a)	URL (Feb 25, 2025)	Guette (2017b)	×
Coworking establishments	Location of WeWork and Regus coworking spaces	Ahlfeldt et al. (2025b)	DOI (Feb 27, 2025)*		×
Commercial office stock	Real estate investment trust (REITS) investments of 2014	Originator: SNL Real Estate Research (2014); transformed version provided by Ahlfeldt et al. (2025b)	proprietary data (Aug 25, 2015)	SNL Real Estate Research (2014)	× proprietary data
Population grid	Gridded Population of the World Dataset Version 4.10	NASA Socioeconomic Data and Applications Center (SEDAC)	DOI (Feb 25, 2025)	Center for International Earth Science Information Network (2017)	× manual GIS processing
Metro population	World Urbanization Prospects: The 2014 Revision	United Nations Population Division (2014a)	URL (Feb 25, 2025)	United Nations Population Division (2014b)	× ^α
Subway dataset	Subway census (v1)	Gonzalez-Navarro and Turner (2019)	DOI (Feb 25, 2025)	Gonzalez-Navarro and Turner (2018)	× [†]

Notes: * Data was collected for scraped/collected by the authors for this paper; see documentation.

^α We add populations for Swedish metropolitan areas that were missing in the source (documented for the respective cities in Ahlfeldt et al., 2020).

[†] We perform manual corrections of small deficiencies in the dataset (documented for the respective cities in Ahlfeldt et al., 2020).

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

```

1 begin
2   for each city do
3     for  $i \leq 100K$  do
4       draw a developable random location  $l_i$  within the city
5       compute employment  $e_i$  within a disk of radius 750 metres around  $l_i$ 
6       adjust  $e_i$  for the share of undevelopable cells in the disk
7     Compute the 99.5th percentile of the distribution of the 100K  $e_i$ 
8     for all plants  $p_i$  in our city do
9       compute employment  $\tilde{e}_i$  within a disk of radius 750 metres around  $p_i$ 
10      if  $\tilde{e}_i$  above the 99.5th percentile of  $e_i$  distribution then
11        mark  $p_i$  as cluster point
12    map all cluster points to grid cells; grid cells with at least one cluster point
    are called ‘cluster grid cells’
13    mark all cluster grid cells as unprocessed, sort them in decreasing order of
    cluster point weights
14    while there are unprocessed cluster grid cells do
15      pick the unprocessed cluster grid cells with the largest weight, give it a
      cluster index
16      find all connected cells, give them the same cluster index, mark them as
      processed
17    rank clusters  $c_j$  by decreasing employment
18    for all clusters  $c_j$  in the city do
19      if employment in  $c_j$  less than  $\mathcal{T} = 5\%$  of largest cluster  $c_1$  in city then
20        drop cluster
21      else
22        Assign prime location status
23    while Prime location identifier changes do
24      generate pairs of nearest prime locations
25      if centroid-distance  $\mathcal{D} \leq 2.5\text{km}$  and border-distance  $\mathcal{D} \leq 1\text{km}$  then
26        merge smaller to larger prime location
27      else
28        do nothing
29    for each prime location do
30      Generate orthogonal convex hull
31      if vertical or horizontal extent  $\geq 5\text{km}$  and there is a thin waist within 2.5km of
      the geographic centroid then
32        Split at the waist
33      else
34        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.

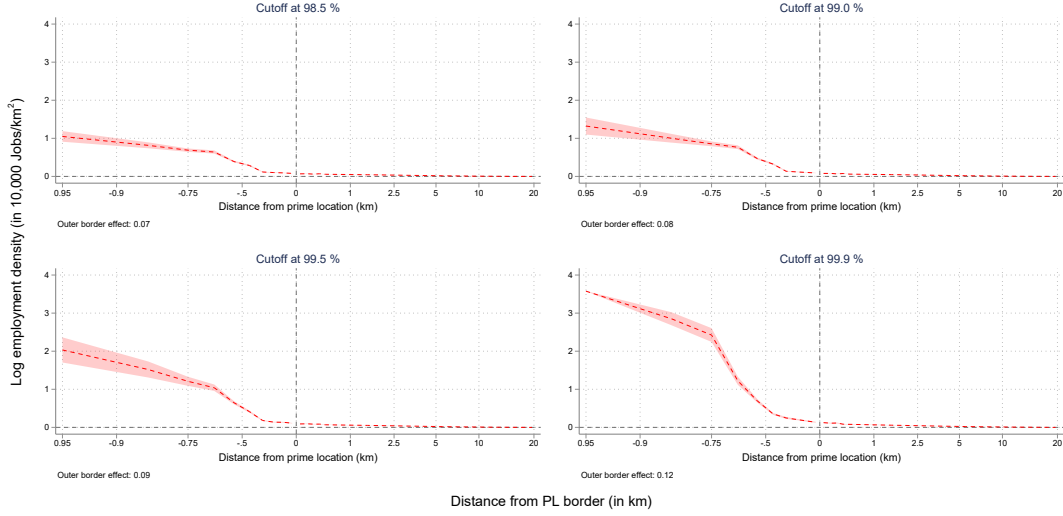
Low cutoffs such as the 98.5th and 99th percentile of the distribution of employment densities across developable cells identify areas that are not dense enough (see the first rows of Figure B.1.1). Too high cutoffs such as the 99.9th percentile result in a ‘spatial pre-trend’, i.e., density increases 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 location (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 areas

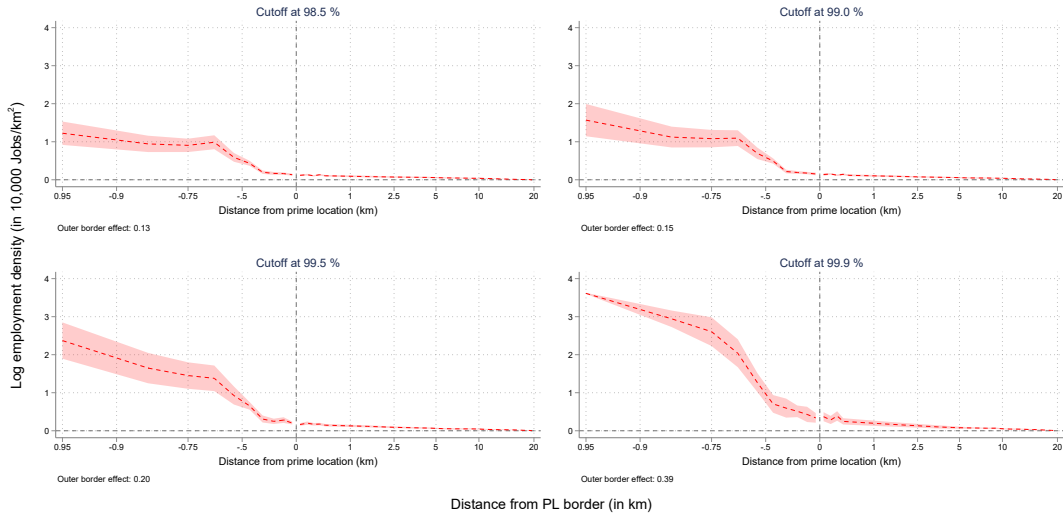
Prime locations and historic manufacturing shares In Table B.2.1, we regress the number of prime locations and their share at total MSA employment against the 1940 share of manufacturing establishments at manufacturing and services employment. We control for city size since the historic manufacturing 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 (0.0045)	-0.009 (0.0037)	-0.183 (0.0749)	-0.052 (0.0301)	-0.048 (0.0276)	0.528 (0.1389)
Log employment		0.689 (0.0913)	1.990 (0.7150)		0.836 (0.2161)	2.230 (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," available 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 (0.0789)	-0.048 (0.0512)	-1.105 (0.6124)	0.471 (0.3953)	0.143 (0.4070)	-0.946 (1.2474)
Log employment		0.743 (0.1003)	1.891 (0.8823)		0.856 (0.2385)	3.002 (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	-0.020 (0.0267)	-0.050 (0.0237)	-0.431 (0.4089)	0.613 (0.1524)	0.582 (0.1527)	1.437 (0.4932)
Log employment		0.878 (0.1228)	1.872 (0.8310)		0.900 (0.2960)	2.536 (1.5283)
MSAs	All	All	Emp. $\geq$ 1M	All	All	Emp. $\geq$ 1M
Observations	220	220	36	220	220	36
R^2	.00121	.477	.249	.0954	.138	.263

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 (0.0044)		0.086 (0.0129)		-0.013 (0.0048)	
HHI, outside PL	-0.135 (0.6644)		0.490 (0.6941)		1.091 (2.5684)	
PL rank		0.035 (0.0185)		0.123 (0.0990)		0.026 (0.0094)
MSA effects	-	Yes	-	Yes	-	Yes
MSAs	All	All	Emp. < 1M	Emp. < 1M	Emp. ≥ 1M	Emp. ≥ 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 (0.0264)	0.127 (0.0206)	0.224 (0.0225)	0.199 (0.0156)	0.050 (0.0085)	0.058 (0.0059)
Number of prime locations	-0.021 (0.0249)		-0.042 (0.0167)		0.009 (0.0055)	
PL share at total employment (%)		0.005 (0.0043)		-0.003 (0.0033)		-0.003 (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 Single-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 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

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 (0.0013)	-0.064 (0.0014)	0.001 (0.0003)	-0.005 (0.0003)	0.008 (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 (0.0026)	-0.072 (0.0027)	0.005 (0.0010)	-0.002 (0.0007)	0.015 (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 (0.0053)	-0.079 (0.0056)	0.013 (0.0012)	0.002 (0.0011)	0.021 (0.0013)
Observations	1540	1539	1538	1491	1538
R^2	.703	.705	.47	.618	.442
Panel (d): Prime location rank within MSA ≥ 2 (large MSAs)					
Distance from PL (km)	-0.068 (0.0029)	-0.070 (0.0031)	0.002 (0.0011)	-0.004 (0.0008)	0.012 (0.0010)
Observations	3720	3720	3717	3585	3715
R^2	.636	.649	.385	.594	.312

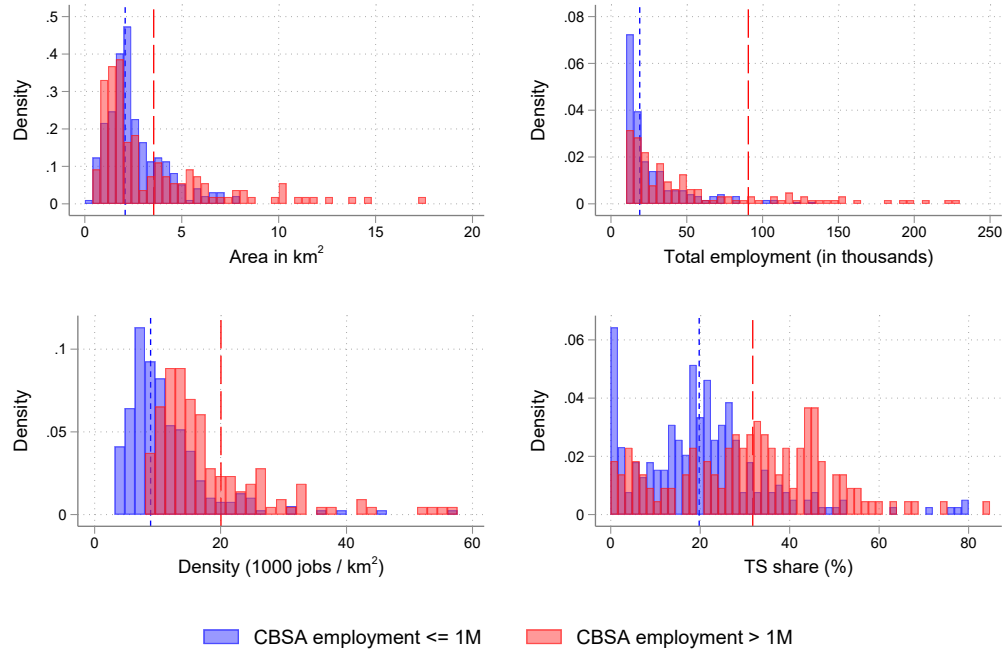
Notes: Unit of observation is 1km-distance from prime location bin by prime location catchment area (observations sharing the same nearest prime location). Bins are aggregated from census block groups and restricted to bins up to 50 km from the nearest prime location. All models include prime location 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, available 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

CBSA	Classification	# PL	Prime locations' share (%) at		
			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 $< 1M$		1.145	0.10%	10.76%	14.85%

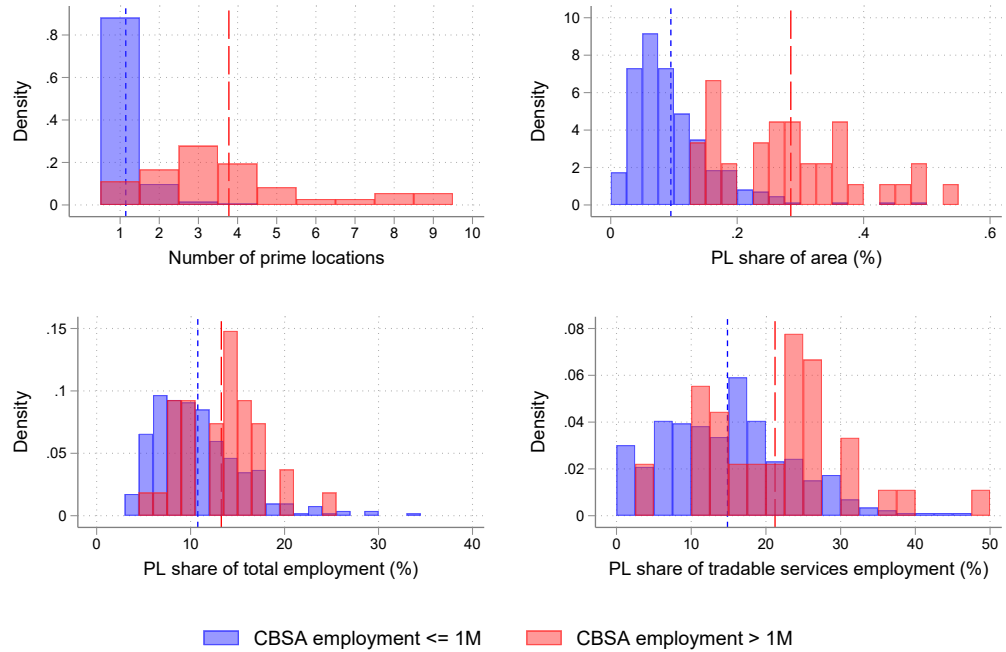
Notes: Agglomerated and dispersed cities are defined by on whether the share of prime locations at total employment is larger or smaller than the median share. We define an MSA as agglomerated if prime locations' share at MSA's tradable services employment is above the median within its size category (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



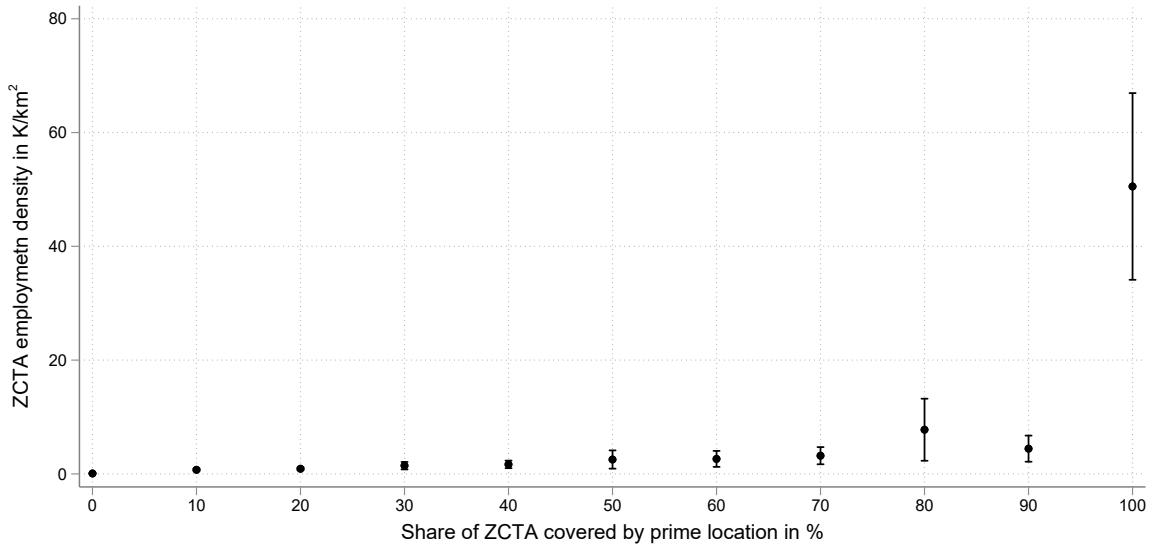
Notes: Unit of observation is prime location. To improve readability, we drop outliers that exceed the following values: Area $\geq 50 km^2$, employment $\geq 250k$, employment density $\geq 100k$.

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%.

(GIS) and merge the CBP zip code employment to the ZCTAs using an official 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.

B.3.2 Summary statistics

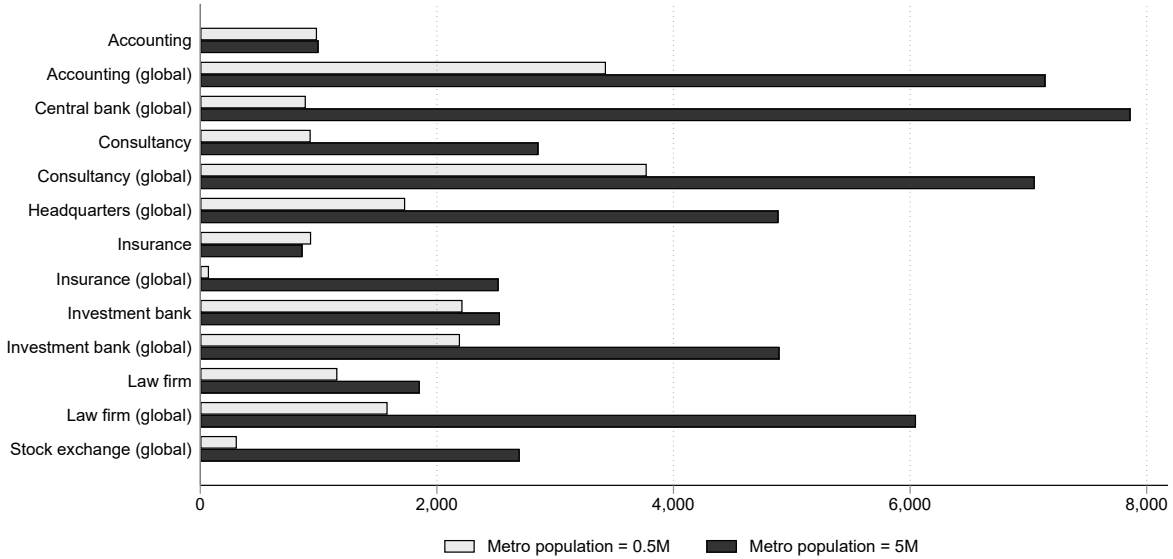
Figure B.3.2 illustrates the distributions of key features of the 286 prime locations we delineate in the Global Cities sample.

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.

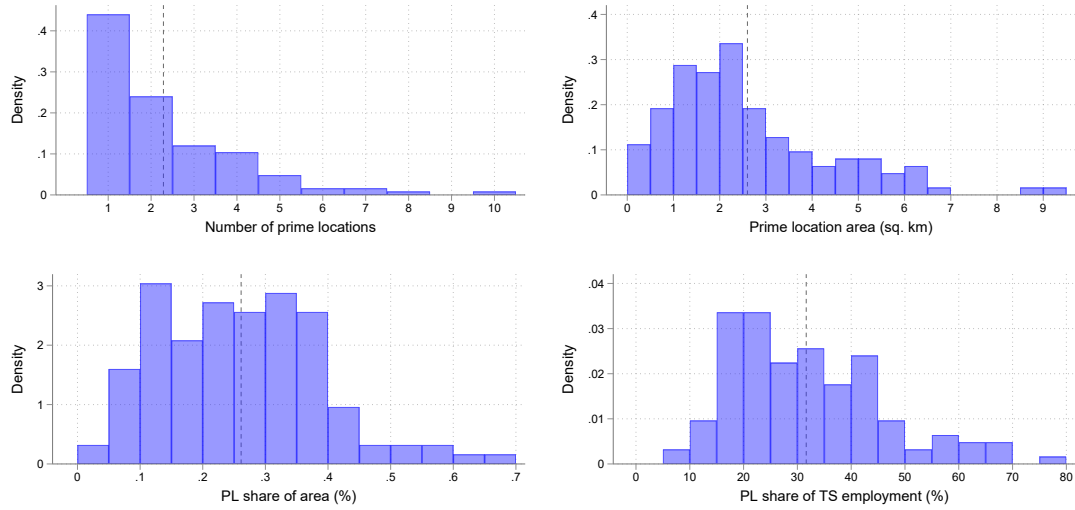
Figure B.3.1: Estimated employment weights by city size: All MSAs



B.3.3 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 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. Col-

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.

umn (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 (0.0038)	0.746 (0.0035)	0.712 (0.0141)	0.824 (0.0130)
Constant	0.001 (0.0000)	0.000 (0.0000)	0.002 (0.0000)	0.002 (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 algorithm, 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	0.053	0.029	0.034	0.038	0.026
	S.E.	0.004	0.015	0.003	0.004	0.007	0.005
	R2	0.259	0.305	0.776	0.513	0.322	0.300
SNL-S&P investments	Coeff.	0.115	0.159	0.075	0.071	0.052	0.037
	S.E.	0.038	0.071	0.025	0.011	0.012	0.001
	R2	0.416	0.306	0.569	0.270	0.101	0.573
Starbucks	Coeff.	0.011	0.084	0.008	0.029	0.075	0.009
	S.E.	0.006	0.022	0.001	0.002	0.013	0.006
	R2	0.052	0.169	0.580	0.397	0.268	0.045
# cities		6	23	3	34	54	5

Notes: ^aIncluding Middle East. ^bNorth America. ^cSouth 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	PL's share at area	PL's share at TS employment
Africa & Middle East	6	2.00	0.33%	29.06%
Asia	23	2.00	0.25%	37.81%
Australia	3	1.33	0.34%	36.61%
Europe	34	1.35	0.34%	31.34%
North America	54	3.17	0.20%	27.59%
South America	5	1.40	0.33%	49.26%

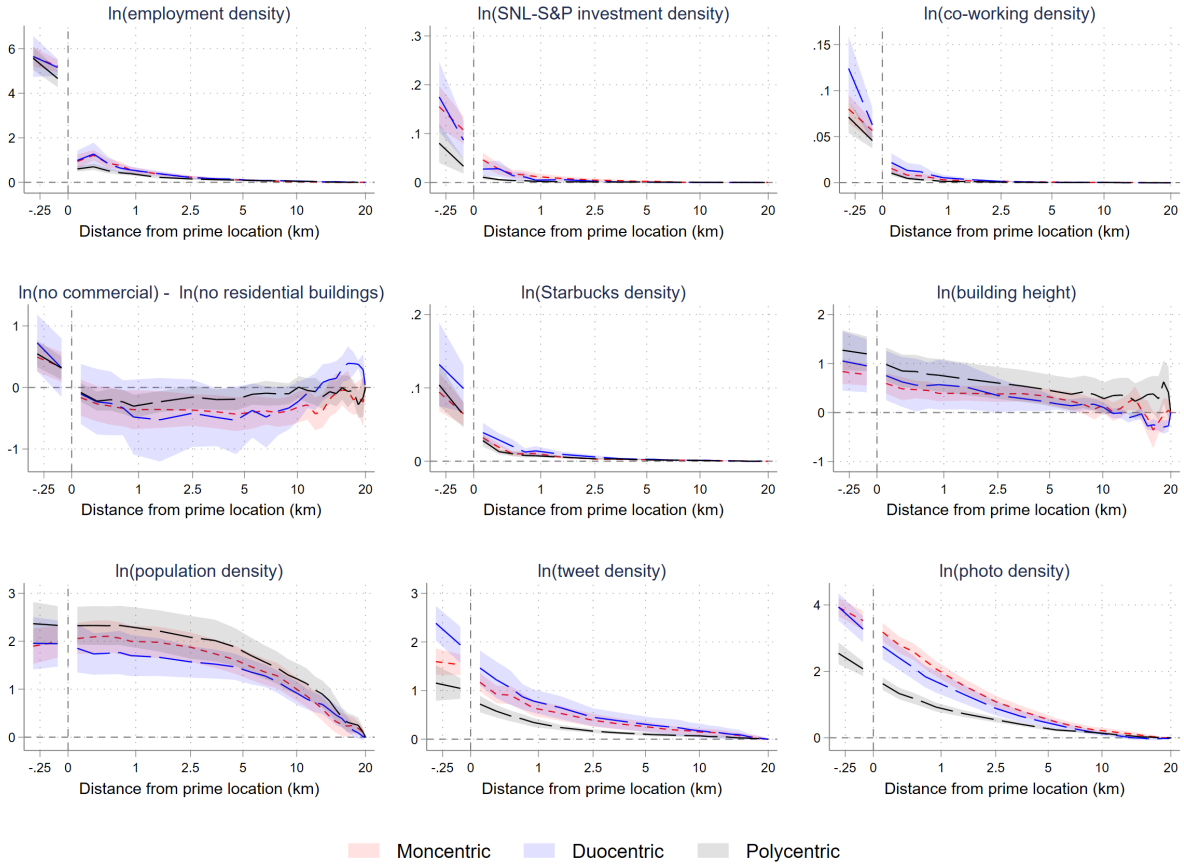
Notes: Prime location (PL) share at tradable services (TS) employment is the share of prime locations at total predicted tradable services employment.

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 and Data References

- Ahlfeldt, Gabriel M. and Daniel McMillen, "Replication Data for: "Tall Buildings and Land Values: Height and Construction Cost Elasticities in Chicago, 1870 – 2010"," Technical Report, Distributed via Havard Dataverse; DOI 2017.
- 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.
- and —, "Tall Buildings and Land Values: Height and Construction Cost Elasticities in Chicago, 1870–2010," *The Review of Economics and Statistics*, 12 2018, 100 (5), 861–875.
- Ahlfeldt, Gabriel M, Thilo Albers, and Kristian Behrens, "Prime locations," *CEPR Discussion Paper No. 15470.*, 2020.

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 $d > 1\text{km}$, 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 ($d > 0$). 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.

—, —, and —, "Prime locations toolkit 1.0.0, 2025-01," Technical Report, distributed by Gabriel Ahlfeldt via GitHub; [URL](#) (last accessed: Feb 25, 2025) 2025.

Ahlfeldt, Gabriel, Thilo Albers, and Kristian Behrens, "Replication Data for "Prime Locations"," Technical Report 2025.

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

Behnam, Vafa, "ICPSR_FIPS: Correspondence between ICPSR and FIPS county codes," Technical Report, Github repository 2020.

Census Bureau, "Census Core-Based Statistical Area (CBSA) to Federal Information Processing Series (FIPS) County Crosswalk," Technical Report, Hosted by NBER 2023.

Center for International Earth Science Information Network, "Gridded Population of the World, Version 4 (GPWv4): Basic Demographic Characteristics, Revision 10. Palisades, NY: NASA Socioe-

- conomic Data and Applications Center (SEDAC)," Technical Report 2017.
- Chen, J. J. and Joe Wasserman**, "ZCTA: ZIP Code Tabulation Area (ZCTA) Relationship files," Technical Report, GitHub 2024.
- D., Van De Kerchove R. Daems D. De Keersmaecker W. Brockmann C. Kirches G. Wevers J. Cartus O. Santoro M. Fritz S. Lesiv M. Herold M. Tsendbazar N.E. Xu P. Ramoino F. Arino O. Zanaga**, "ESA WorldCover 10 m 2021 v200," Technical Report, European Space Agency 2022.
- Danielson, Jeffrey J and Dean B Gesch**, "Global multi-resolution terrain elevation data 2010 (GMTED2010)," Technical Report, US Geological Survey 2011.
- Earth Resources Observation And Science Center**, "Global Land Cover Characterization (GLCC)," Technical Report, distributed by the United States Geological Survey; DOI 2017.
- , "Global Multi-resolution Terrain Elevation Data 2010 (GMTED2010)," Technical Report, Distributed by the United States Geological Survey; DOI 2017.
- European Space Agency**, "ESA WorldCover 10 m 2021 v200," Technical Report, European Space Agency; DOI 2022.
- Feng, Min, Joseph O. Sexton, Saurabh Channan, and John R. Townshend**, "A global, high-resolution (30-m) inland water body dataset for 2000: first results of a topographic-spectral classification algorithm," *International Journal of Digital Earth*, 2016, 9 (2), 113–133.
- Fischer, Erica**, "The Geotaggers' World Atlas," Technical Report 2016.
- Gonzalez-Navarro, Marco and Matthew A. Turner**, "Subways and urban growth: Evidence from earth," *Journal of Urban Economics*, November 2018, 108, 85–106.
- and —, "Replication Data for: Subways and Urban Growth: Evidence From Earth," Technical Report, distributed via Harvard Dataverse; DOI 2019.
- Guette, Kaylee**, "All Starbucks Locations In The World," Technical Report, Data distributed via Socrata Open Data Repository; url (last accessed: Feb 25, 2025) 2017.
- , "All Starbucks Locations In The World," Technical Report, Socrata Open Data Repository 2017.
- Gyourko, Joseph, Jonathan S. Hartley, and Jacob Krimmel**, "The local residential land use regulatory environment across U.S. housing markets: Evidence from a new Wharton index," *Journal of Urban Economics*, 2021, 124, 103337.
- Hansen, Matthew C, Ruth S DeFries, John RG Townshend, and Rob Sohlberg**, "Global land cover classification at 1 km spatial resolution using a classification tree approach," *International Journal of Remote Sensing*, 2000, 21 (6-7), 1331–1364.
- NASA Socioeconomic Data and Applications Center (SEDAC)**, "Gridded Population of the World, Version 4 (GPWv4): Basic Demographic Characteristics, Revision 10. Palisades, NY: NASA Socioeconomic Data and Applications Center (SEDAC)," Technical Report, Distributed via Basic Demographic Characteristics, Revision 10. Palisades, NY: NASA Socioeconomic Data and Applications Center (SEDAC); DOI 2017.
- National Historical Geographic Information System**, "IPUMS National Historical Geographic Information System: Version 19.0 .," Technical Report, IPUMS NHGIS, Minneapolis, MN: IPUMS; DOI 2024.
- PricewaterhouseCoopers**, "UK Economic Outlook (November 2009) – Which are the largest city economies in the world and how might this change by 2025?," 2009.
- Schroeder, David Van Riper Katherine Knowles Tracy Kugler Finn Roberts Steven Manson Jonathan and Steven Ruggles**, "IPUMS National Historical Geographic Information System: Version 19.0 .," Technical Report, Minneapolis, MN: IPUMS. 2024.
- SNL Real Estate Research**, "Real estate investment trust (REITS) investments," Technical Report, SNL (now part of S&P) 2014.

- Trujillo, Jesus Leal and Joseph Parilla**, *Redefining global cities*, The Brookings Institution, 2016.
- United Nations Population Division**, “World Urbanization Prospects: The 2014 Revision, CD-ROM Edition,” Technical Report, Data distributed via London Data Store; [URL](#) (last accessed Feb 25, 2025) 2014.
- , “World Urbanization Prospects: The 2014 Revision, CD-ROM Edition,” Technical Report 2014.
- United States Census Bureau**, “County Business Patterns: 2014,” Technical Report, Distributed via United States Census Bureau; [URL](#) (last accessed: Feb 25, 2025) 2016.
- , “County Business Patterns: 2014,” Technical Report, United States Census Bureau 2016.
- United States Geological Survey**, “Global Forest Cover Change (GFCC) Water Cover 2000 Global 30 m (GFCC30WC v001),” Technical Report, distributed by the United States Geological Survey; [URL](#) (last accessed: Feb 25, 2025) 2016.
- Walls, Donald W**, “National Establishment Time-Series Database©: Presentation,” in “2007 Kauffman Symposium on Entrepreneurship and Innovation Data” 2007.
- Wharton School**, “Wharton Residential Land UseRegulatory Index,” Technical Report, Distributed by Wharton School, University of Pennsylvania ; [URL](#) (last accessed: Feb 25, 2025) 2020.
- Zillow Research**, “Housing Data,” Technical Report, Data distributed via Zillow’s official website; [URL](#) (data retrieved Oct 15, 2024) 2024.
- , “Housing Data,” Company Website 2024.