A CoStar Data Validation

As a first attempt at validating the coverage and content of the CoStar data, we compare them to three alternative data sources. First, we consult CBRE, a global real estate services company that collects data on retail lease rates over time. CBRE makes data on rents available at aggregate levels (i.e., citywide and sub-municipal markets), and we compare CoStar and CBRE rents over time in Appendix Figure 1. While both datasets use "gross asking rents," the levels are different; this is likely due to different sources or definitions of what is included in the asking rent figure. However, we are reassured by the similar trends and orders of magnitude across the two datasets over time.

Second, for two of our cities, we confirm that the coverage of CoStar is comprehensive, when compared to the number of establishments reported in public Census products. For example, in New York and Los Angeles, CoStar (as of early 2020) tracks 156,839 and 256,846 commercial leases, respectively. These figures are for all types of commercial, including those beyond retail. The Census' publicly available 2018 County Business Patterns aggregated data report approximately 214,000 and 260,000 customer-facing establishments respectively in New York and Los Angeles. Since these metrics are tracking slightly different phenomena (e.g., some establishments may not have leases or the timing of the aggregate establishments may not line up with the lease terms), they will not be identical; however, we are reassured by the consistent orders of magnitude.

Finally, we compare the number of leases in CoStar data with the number of establishments in each county, as measured by DataAxle.[15] Appendix Table 1 shows that while the number of leases (unsurprisingly) represents only a very small share of all of the establishments in each city, the share is relatively consistent across cities. Boston and Chicago are slightly less covered than the other cities. When we track the lease coverage over time, we also see that the rapid growth in CoStar leases seems to slow down and stabilize in the late 2000s (see Appendix Figure 2). This trend, which is consistent across all of our cities, suggests that the CoStar data is likely the most reliable from around 2007.

---

[15]We access these data from via a Wharton Research Data subscription provided by Harvard University.

Appendix Figure 1: CBRE and CoStar Rent per Square Foot, 2022 Dollars


Note: This figure shows median CoStar retail rent per square foot (largely asking rent, but in some cases effective or starting rent) in 2022 dollars in orange and CBRE mean retail gross asking rent per square foot in purple. CBRE and CoStar use different underlying samples to create median/mean values.

Table 1: Comparison of CoStar Leases and InfoUSA Establishments

| Market | Average Annual Totals | | | Lease Share of Estabs. | |
	Estabs.	Consumer-Facing Estabs.	New Leases	Consumer-Facing	All
Boston	70,651	27,861	940	0.034	0.013
Chicago	210,231	47,109	2,588	0.055	0.012
Houston	80,976	33,968	2,148	0.063	0.027
Los Angeles	145,600	63,147	3,296	0.052	0.023
Miami	40,967	19,731	815	0.041	0.020
New York	123,201	50,730	2,599	0.051	0.021
Washington	71,130	25,779	1,562	0.061	0.022

Notes: This table reports CoStar lease data and InfoUSA establishment data for 2005 to 2021. Displaying the average number of establishments per year (for InfoUSA data), or the average total number of new leases per year (for CoStar data). Consumer-facing establishments are all establishments in NAICS sectors 44-45, 71, 72, 81 and 311811 (retail bakeries).

Appendix Figure 2: New Leases as a Share of NAICS 44-45 Establishments by Year and Market
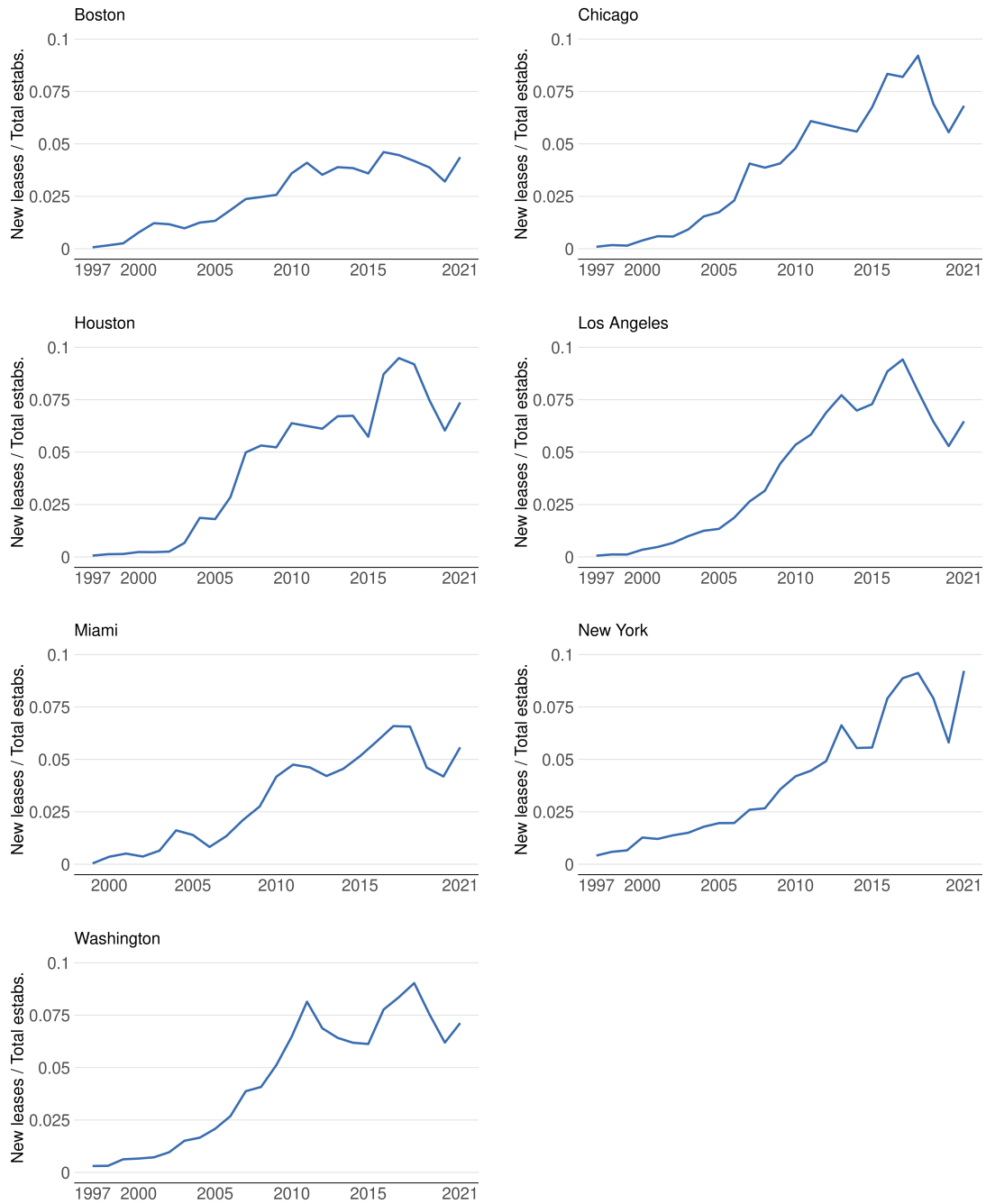
Note: This figure uses CoStar lease data and DataAxle/InfoUSA establishment data. For each market, each sub-figure reports the number of new CoStar leases in the year on the horizontal axis divided by the number of establishments in NAICS codes 44-45 (retail) in that same year.

B CoStar Data Orientation

We now present summary statistics documenting retail leases from the CoStar data. First, we note that other than square footage, the coverage of information on rents and lease terms is not complete. In Appendix Figure 3 we report the share of observations by year that are missing data on the square footage of the lease, the rent on the lease, the lease location and the lease term (length). CoStar does not have complete coverage for all variables because it relies on self reports from brokers. Brokers are particularly hesitant to share rent and lease length, as it may pose a risk of losing clients to competing brokers (information on square footage does not hold a similar premium). Coverage improves for all of our markets since the mid-2000s.

In addition, we track the spatial expansion of the CoStar coverage over time by regressing the lease-level distance to City Hall on time (years) for each market. These estimates are plotted in Figure 4 . Again we see a stabilization in the average distance across leases after 2005, following increases for most of the markets in earlier years (with the exception of Los Angeles).

Altogether, these patterns, along with the comparisons to establishment counts above, indicate that around 2005-2007 CoStar's coverage becomes closer to the near-universe of leases.

We also assess the CoStar data and its coverage with respect to the main variables of interest. Since the key contribution of the CoStar data is the information on rents, we divide the sample of leases into those with the (asking) rent field populated and those with that field missing. We show summary statistics in Table 2. For each market, we report the number of leases and mean measures of property characteristics based on whether we observe gross asking rent per square foot. While the mean lease start year is similar regardless of rent information status, leases with rents populated tend to record lower leased square footage, fewer months on the market, and shorter term lengths.

Appendix Figure 3: Missing Data Shares by Year and Variable
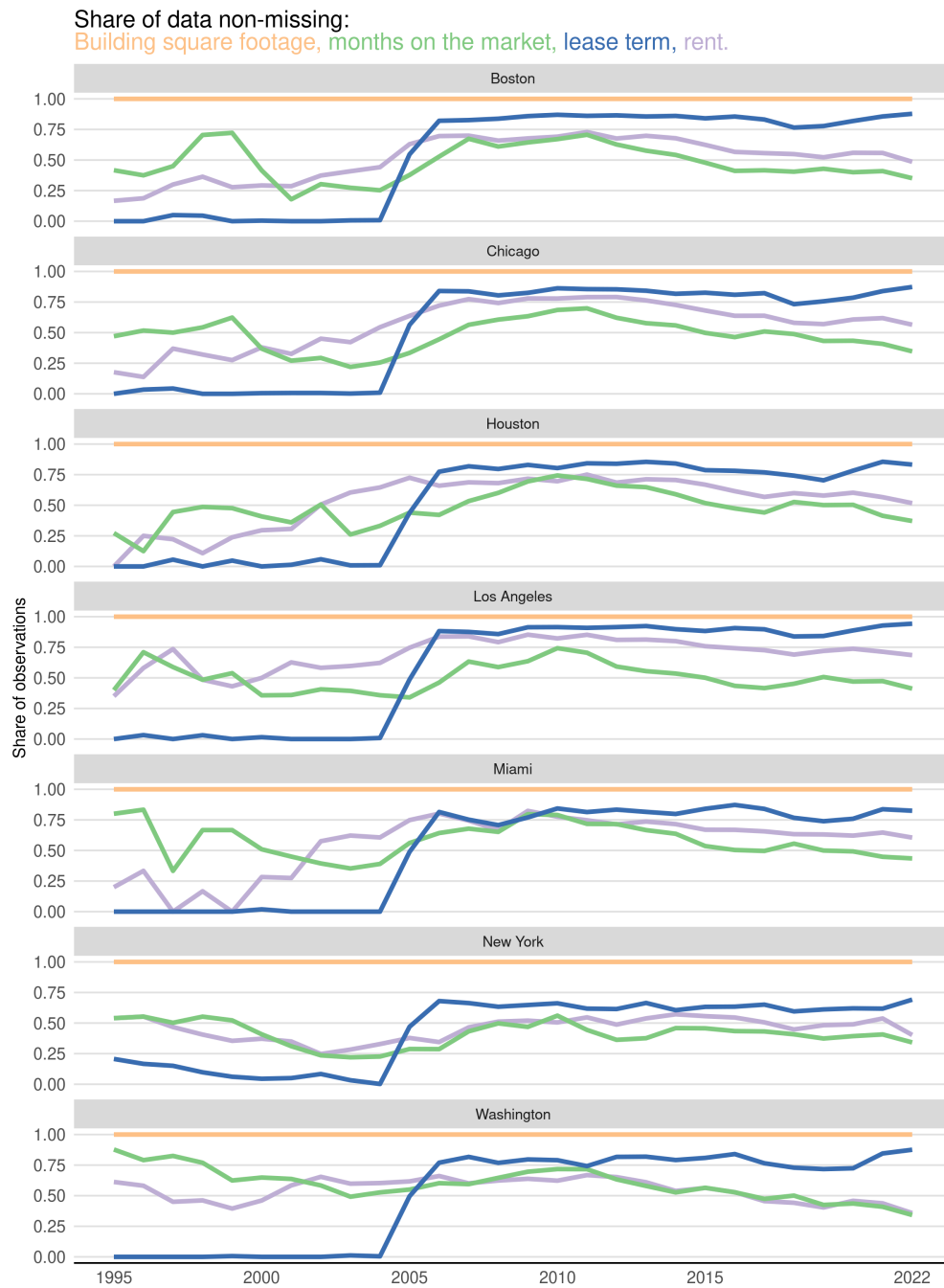
Note: This figure uses CoStar lease data and reports the share of leases with missing information by market and year.

Appendix Figure 4: Average Distance to Center Evens Out After Roughly 2006
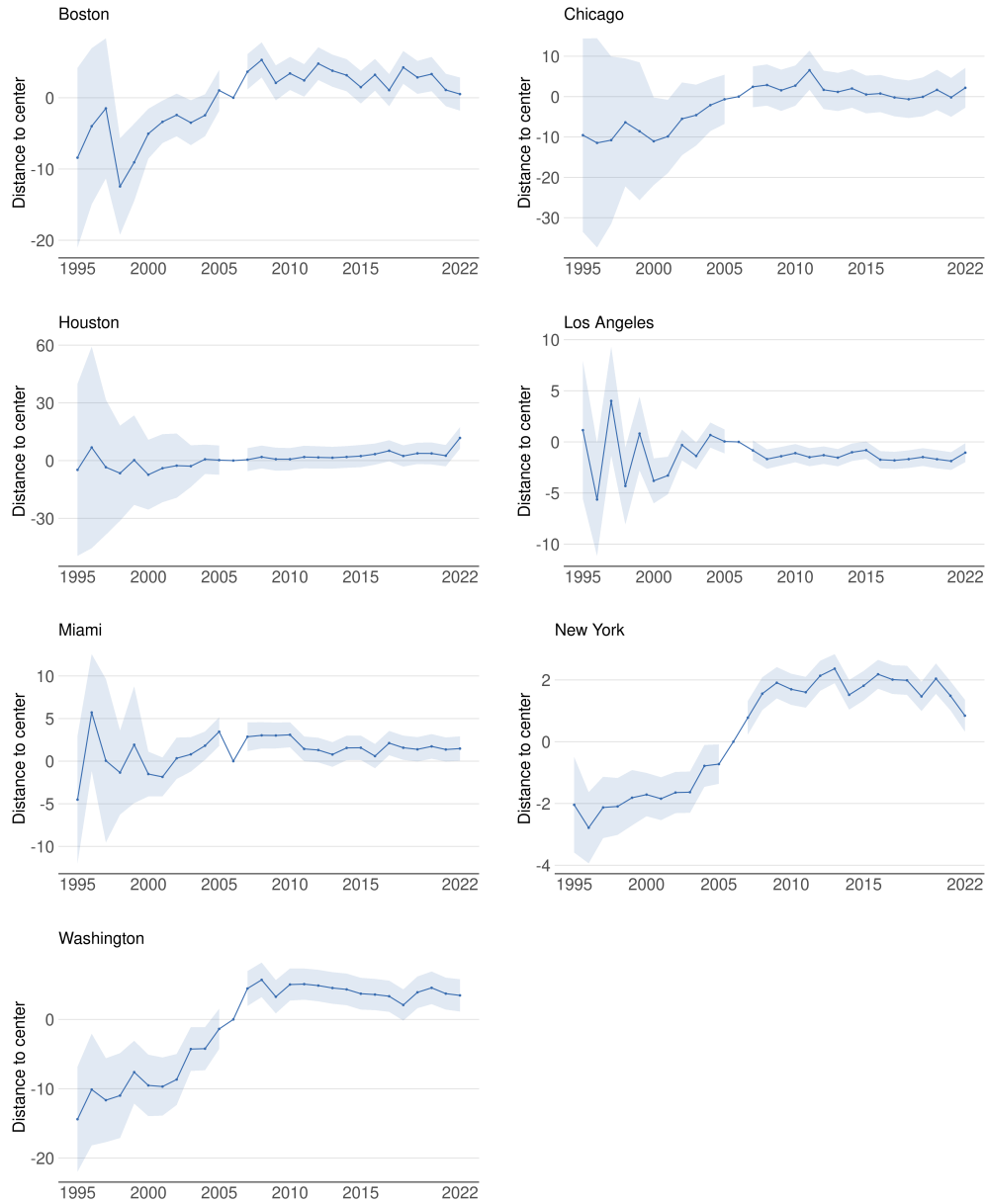
Note: We use geocoded CoStar lease data to calculate the distance between each lease and City Hall of the relevant market. For each market, we then regress this measure of distance on a set year fixed effects, where we omit year 2006. This figure plots the coefficients on these year fixed effects, along with their standard errors (shown by the shading around the line). All values are relative to 2006, which we report as zero. The coefficient of roughly 1 for Boston in 2021 means that the average lease was 1 mile further from Boston's City Hall in 2021 than in 2006. The large negative coefficient for Boston in 1995 means that the average lease in 1995 is almost ten miles closer to City Hall than in 2006.

Table 2: Comparison of Leases With and Without Rent Information

	Boston		Chicago		Houston		Los Angeles	
	w/ rent	w/o rent	w/ rent	w/o rent	w/ rent	w/o rent	w/ rent	w/o rent
Number of Leases	10,510	6,414	31,866	14,717	24,719	13,949	48,626	14,570
Share of Leases	0.62	0.38	0.68	0.32	0.64	0.36	0.77	0.23
Mean Rent per SF, $2022	28.9	.	25.9	.	22.2	.	37.9	.
Lease Start Year	2013.8	2015.1	2014	2015.4	2014.4	2015.5	2014.4	2015.5
Start Year Non-Missing	1	1	1	1	1	1	1	1
Months on Market	15.2	18.2	19.1	20.7	18.7	18.9	13	15.8
Months on Market Non-Missing	0.9	0.73	0.9	0.64	0.87	0.66	0.94	0.72
Lease SF	2,566	5,435	2,799	6,072	3,144	4,725	2,319	4,642
Lease SF Non-Missing	1	1	1	1	1	1	1	1
Term Length	46.5	68	47.5	77.2	48.3	58.9	41.6	54.9
Term Length Non-Missing	0.65	0.28	0.64	0.27	0.66	0.33	0.59	0.3

	Miami		New York		Washington, DC	
	w/ rent	w/o rent	w/ rent	w/o rent	w/ rent	w/o rent
Number of Leases	10,010	4,663	23,740	23,271	15,146	12,963
Share of Leases	0.68	0.32	0.5	0.5	0.54	0.46
Mean Rent per SF, $2022	39.4	.	108.9	.	37.5	.
Lease Start Year	2015.1	2016.1	2014.8	2014.8	2013.7	2015.3
Start Year Non-Missing	1	1	1	1	1	1
Months on Market	15.7	17.7	10.4	12.9	17.3	20.1
Months on Market Non-Missing	0.87	0.66	0.64	0.62	0.85	0.7
Lease SF	2,774	4,500	2,548	3,531	3,183	5,124
Lease SF Non-Missing	1	1	1	1	1	1
Term Length	46.7	58.7	91	95.6	63.3	83.4
Term Length Non-Missing	0.69	0.34	0.49	0.34	0.75	0.31

Notes: This table reports summary statistics for CoStar leases for years 2005 onward. The first row reports the number of leases in each sub-sample and the remaining rows report means. The number of observations in the first two columns may not apply to all calculations in that market; not all leases with rent also contain information on the other variables.

Appendix Figure 5: CoStar Retail and Zillow Residential Rent per Square Foot, 2022 Dollars
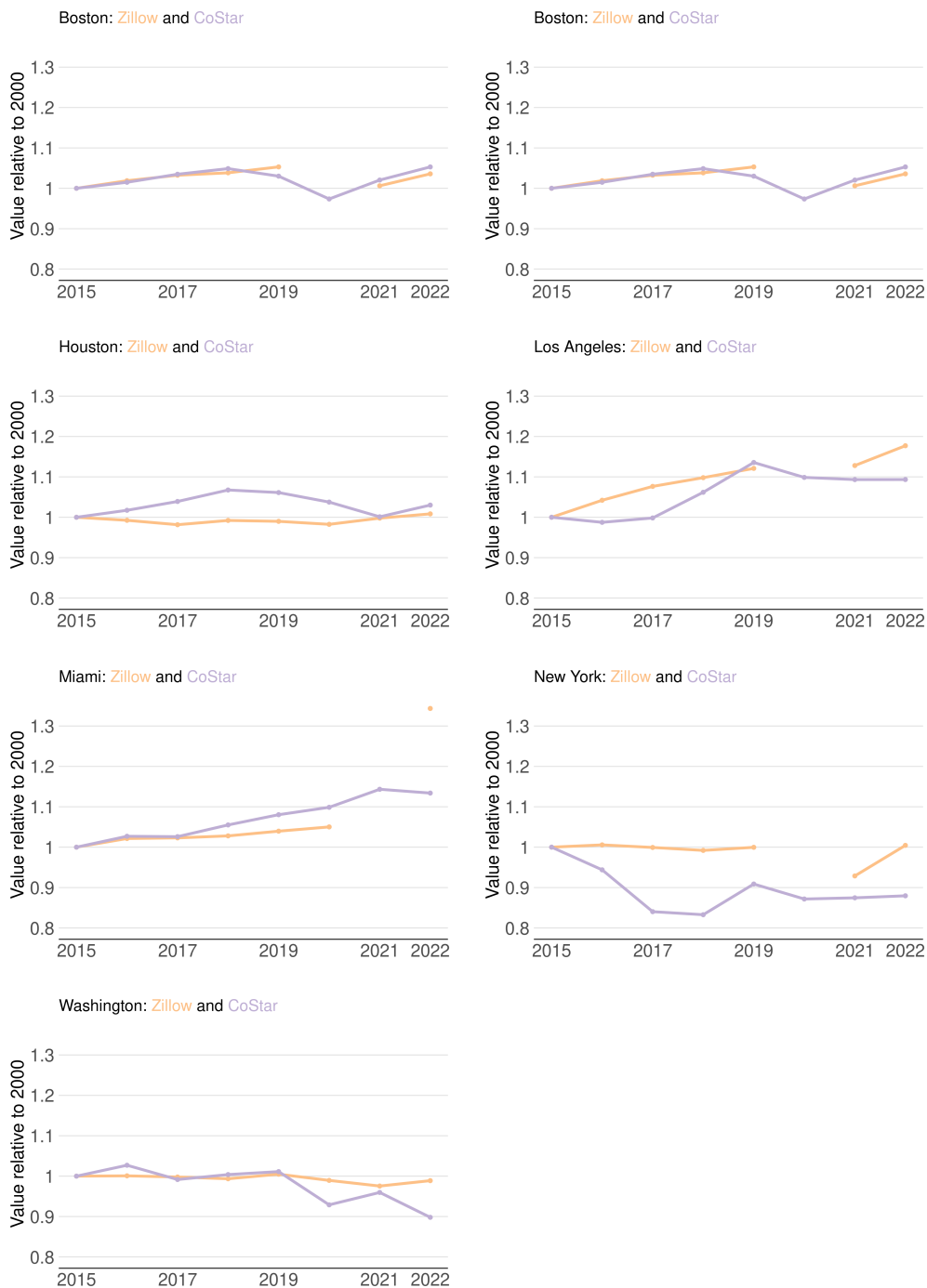
Note: This figure compares CoStar median retail rent per square foot in 2022 dollars (purple) to a Zillow residential rental price index (orange), also in 2022 terms. We normalize both indices to 1 in 2015. Because Zillow does not report a rental index for all markets and years, there are gaps in the orange series.
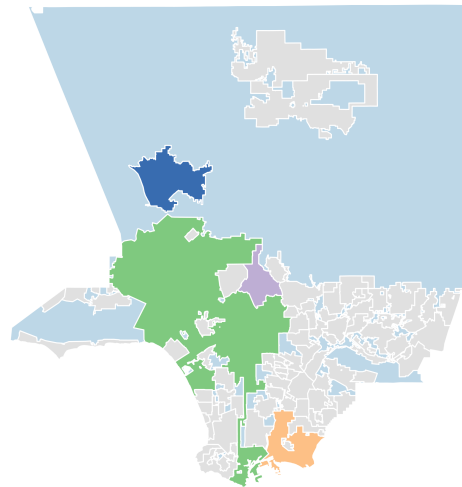
Appendix Figure 6: Maps of New York City and Los Angeles County

(a) New York City

(b) Los Angeles Analysis Areas
Glendale in Purple, Long Beach in Orange, City of Los Angeles in Green, Santa Clarita in Dark Blue, Unincorporated Area in Light Blue, Other Incorporated Areas in Grey


Note: The top figure shows the five boroughs of the city of New York in blue. The bottom figure shows the County of Los Angeles (omitting the offshore islands), with the areas we include in our analysis shown in color (those not included are shown in gray). We include the incorporated municipalities of the City of Los Angeles, Glendale, Santa Clarita and Long Beach, as well as the large unincorporated area shown in light blue.

Sources: City outlines from US Census Bureau place shapefiles, downloaded from NHGIS (Manson et al., 2022). County outline for Los Angeles from Los Angeles City GIS website (City of Los Angeles, 2022).

Appendix Figure 7: Distribution of New Square Footage Leased By Market

Median in dark blue; 25th and 75th percentiles in light blue
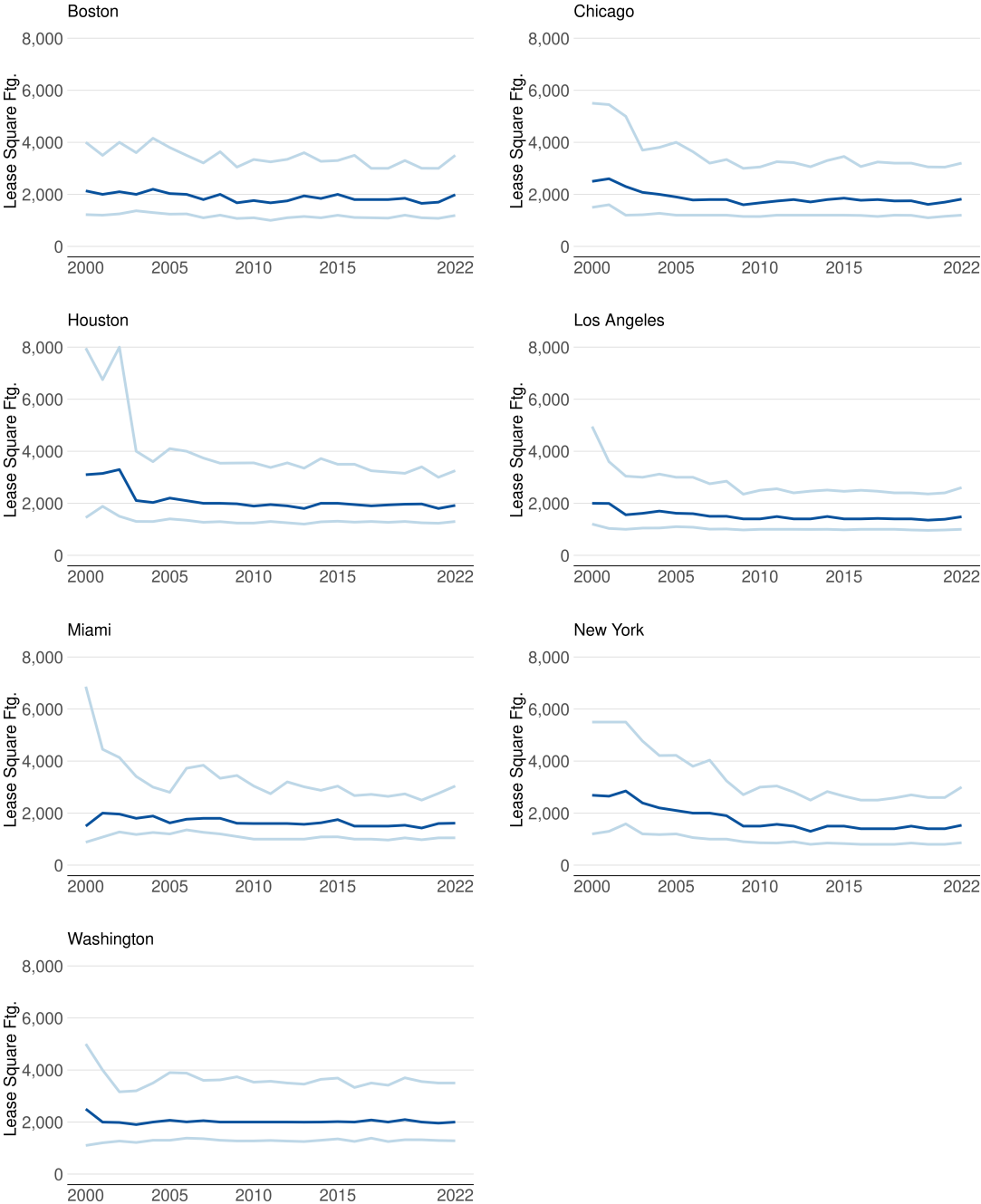
Note: This figure uses CoStar lease data and reports the median (dark blue), 25th percentile and 75th percentile (both in light blue) of leased square footage by market and year.

Appendix Figure 8: Average Length of Lease Roughly Constant 2006 Onward
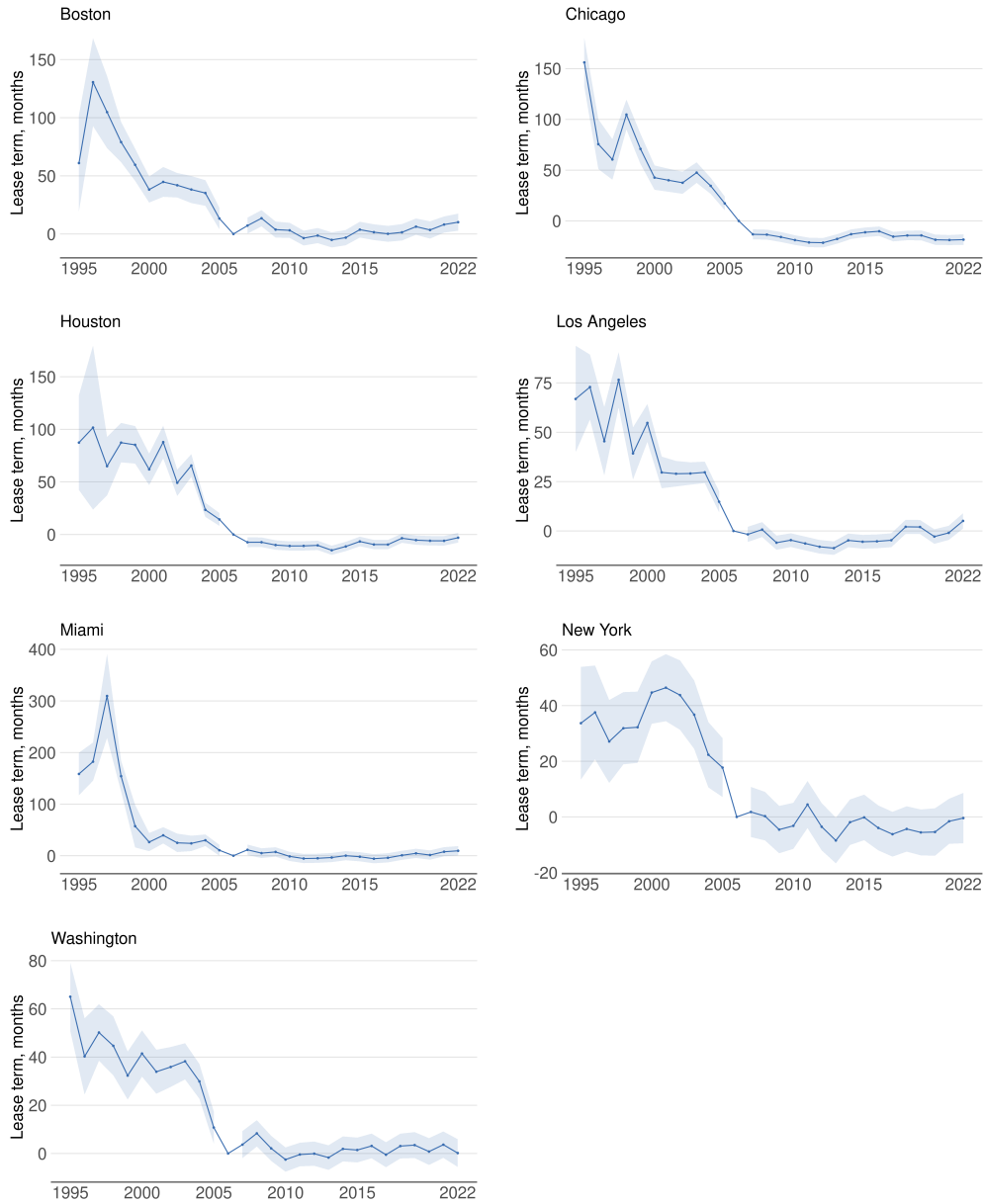
Note: We use geocoded CoStar lease data to regress a lease's term length in months on a set of year fixed effects, where we omit year 2006. This figure plots the coefficients on these year fixed effects, along with their standard errors (shown by the shading around the line). All values are relative to 2006, which we report as zero. The coefficient of roughly zero for Boston in 2017 means that the average 2017 lease had about the same average term length as the average 2006 lease. The large positive coefficients for Boston before 2006 mean that the average CoStar lease recorded before 2006 had much longer terms than the average 2006 lease.

Appendix Figure 9: Average Rent by Year and Market
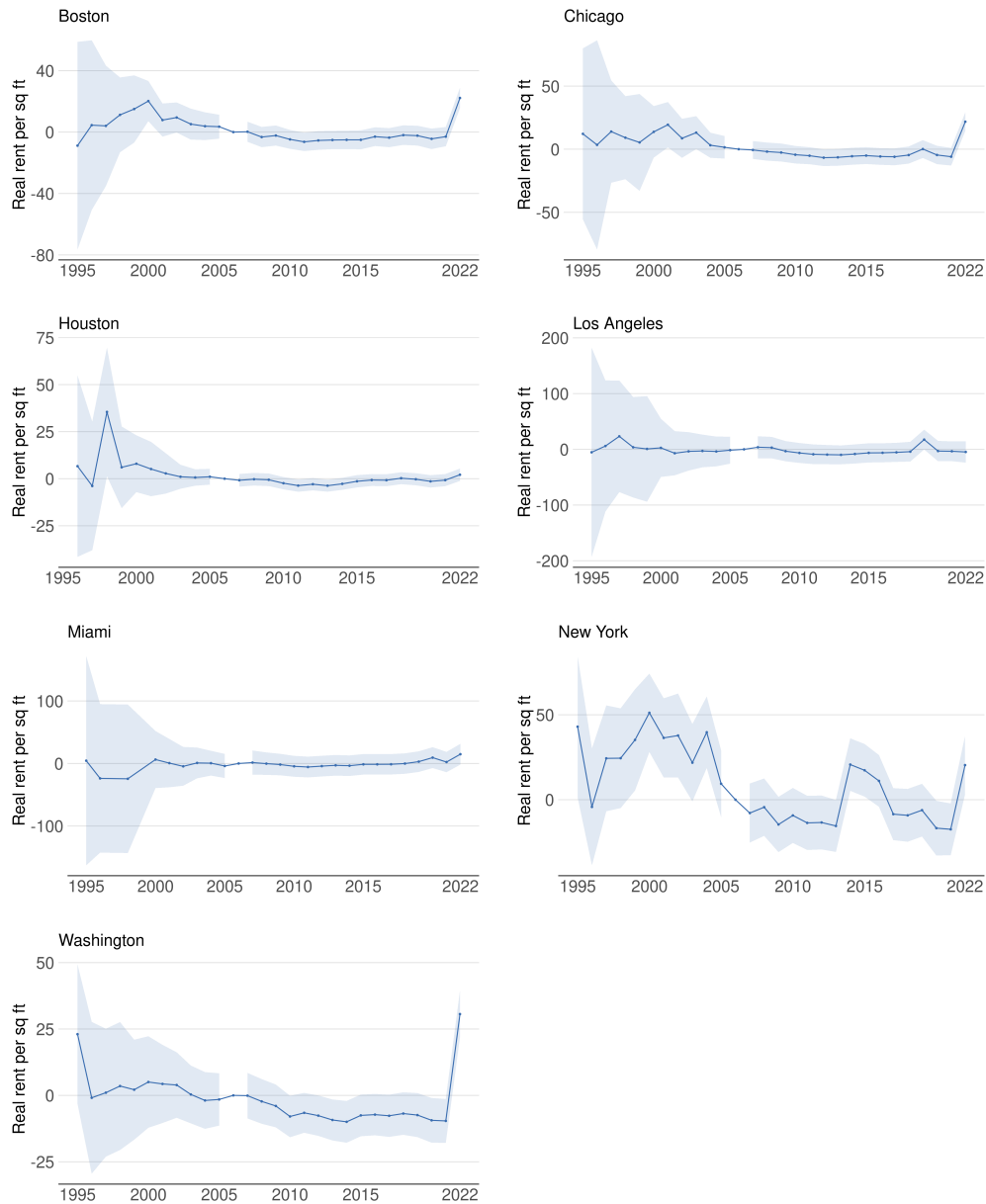
Note: This figure uses CoStar lease data to report the average asking rent per square foot for leases over time. This figure plots the year coefficients from regressions of lease-level rents on year fixed effects, along with their standard errors (shown by the shading around the line), separately for each market. All values are relative to 2006, which we report as zero. The coefficient just below zero for Boston in 2017 means that the average 2017 lease has rent slightly lower than the average average 2006 lease. The larger positive coefficient for Boston around 2000 means that the average CoStar lease recorded in 2000 had higher rent than the average 2006 lease.

Appendix Figure 10: Real Home Prices vs. CoStar Rents, Relative to 2000

Note: This figure shows the median CoStar retail rent per square foot by market and the mean Zillow home price index. We adjust both series for inflation to 2022 dollars, and normalize both series to 1 in 2000.

Appendix Figure 11: CoStar Rents versus CBRE Office and Industrial Rents, Relative to 2001


Note: This figure shows median CoStar retail rent per square foot (purple), and mean CBRE gross asking rent for office (green) and industrial properties (orange), all by market and year. We adjust both series for inflation, and normalize all values to one in 2001 when our data series are complete for all metro areas.
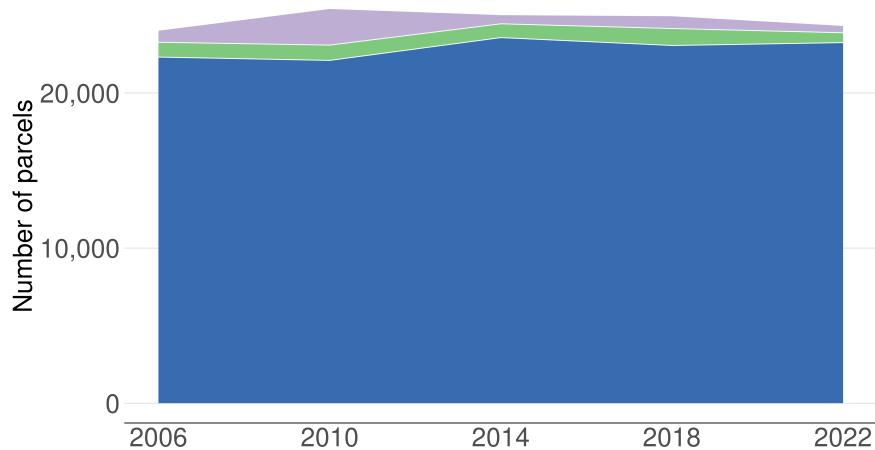
Appendix Figure 12: Most Retail Parcels Do Not Change Zoning Classifications

Parcels that Exit Retail Zoning, Parcels that Enter Retail Zoning, Parcels that Remain Zoned Retail

h (a) New York
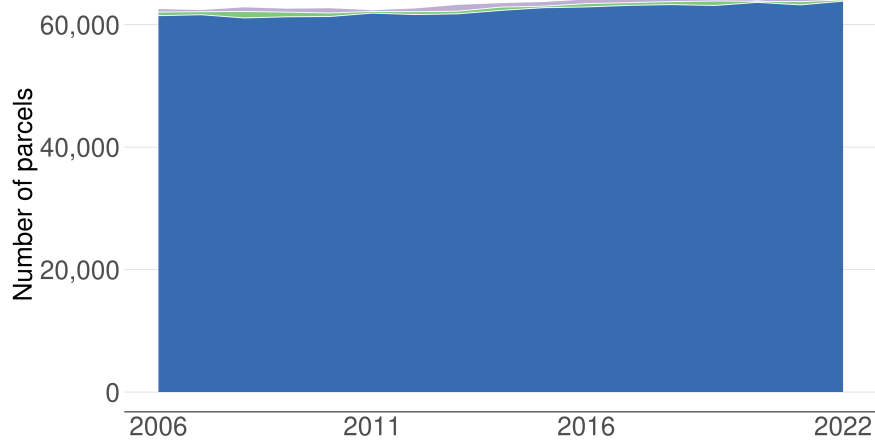
(b) Los Angeles

Note: New York: Land use data are from NYC's PLUTO database. The figure reports the total number of lots that remain zoned as retail (blue), are newly zoned retail (purple), and are converted away from retail zoning (green) in a given year. Los Angeles: Land use data are from the municipal planning departments overseeing the City of Los Angeles, Glendale, Santa Clarita, Long Beach and the unincorporated area of Los Angeles County. The figure reports the total number of lots that remain zoned as retail (blue), are newly zoned retail (purple), and are converted away from retail zoning (green) in a given year. For Los Angeles parcels, "retail" is identified by commercially zoned properties in retail use.

Appendix Figure 13: Distribution of Total Retail Square Footage Concentration, New York
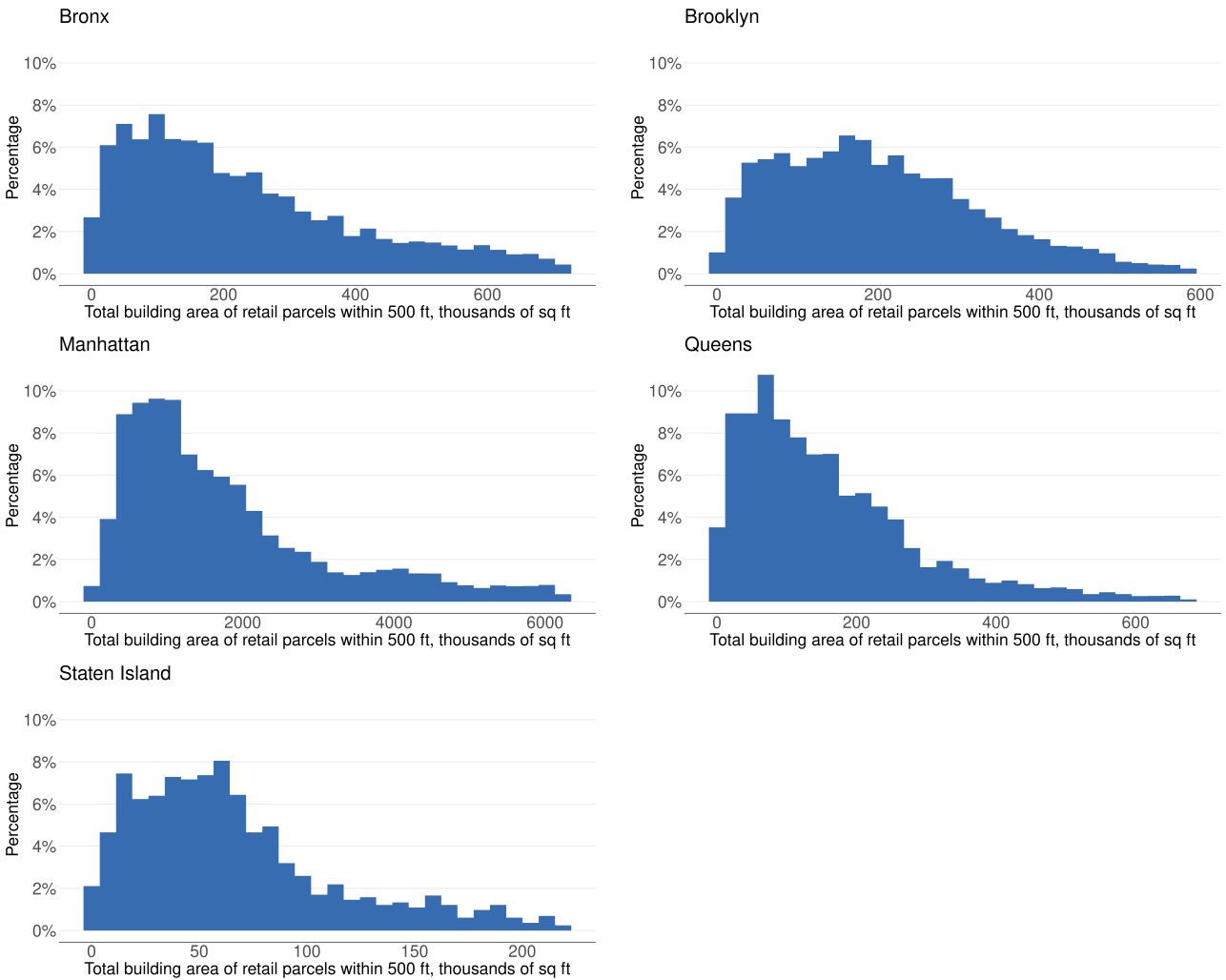
Note: Land use data are from NYC's PLUTO. Retail concentration is measured as the total amount of square footage zoned retail within 500 ft. of a retail-zoned parcel. The figure shows concentration of total square footage of parcels zoned for retail in all five boroughs in 2022. For visibility, we omit the top 5th percentile in each borough. Note that the horizontal axes for Manhattan and Staten Island differ from the other boroughs.

Appendix Figure 14: Distribution of Total Retail Square Footage Concentration, Los Angeles
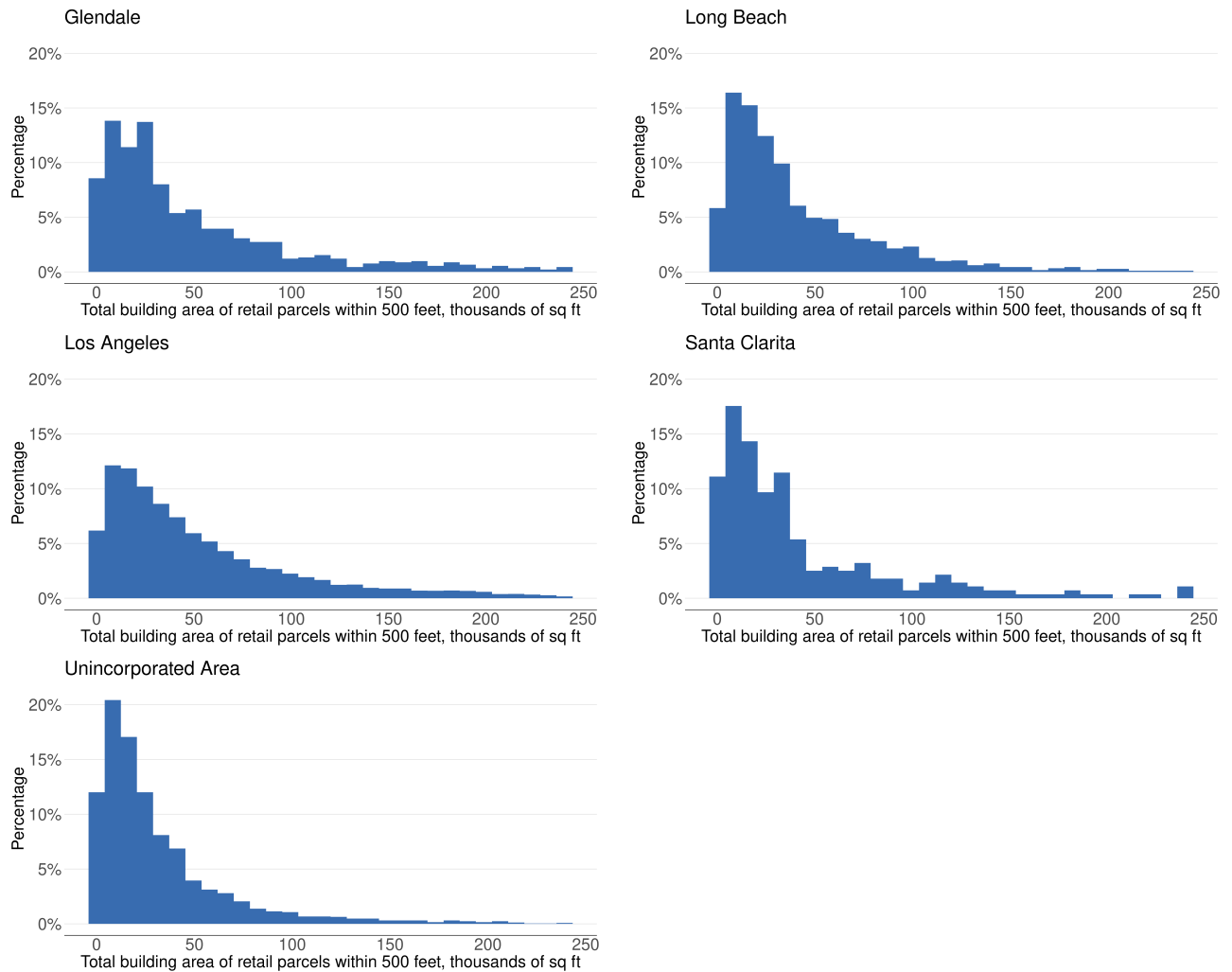
Note: We use only lots in the city of Los Angeles, the unincorporated area of Los Angeles County, the incorporated municipalities of Glendale, Long Beach and Santa Clarita. For Los Angeles parcels, "retail" is identified by commercially zoned properties in retail use. Retail concentration is measured as the total amount of square footage zoned retail within 500 ft. of a retail-zoned parcel. The figure shows concentration of total square footage of parcels zoned for retail in four municipalities and the unincorporated area in 2022. For visibility, we omit the top 5th percentile of values. The distributions are relatively consistent across the cities, with the highest peaks in Long Beach and part of the unincorporated areas (where there are higher concentrations of smaller retail clusters). The City of Los Angeles has the thickest distribution, indicating a wider range of retail clusters and its diversity in land use patterns within the municipality.

Appendix Table 1: Housing and Demographic Summary Statistics by Market

	(1)	(2)	(3)	(4)	(5)	(6)
A. Demographics						
	Share White		Share Black		Share Hispanic	
	2000	2020	2000	2020	2000	2020
Boston	0.807	0.696	0.059	0.074	0.064	0.113
Chicago	0.593	0.524	0.185	0.162	0.164	0.222
Houston	0.48	0.355	0.168	0.168	0.288	0.374
Los Angeles	0.309	0.259	0.096	0.078	0.446	0.483
Miami	0.207	0.135	0.201	0.156	0.573	0.681
New York	0.35	0.319	0.264	0.214	0.27	0.289
Washington	0.553	0.445	0.263	0.248	0.089	0.16
B. Density and Value						
	Population Density		Median Rent		Median Home Value	
	2000	2020	2000	2020	2000	2020
Boston	376	416	773	1,547	212,436	480,659
Chicago	367	382	669	1,166	155,905	250,923
Houston	192	285	595	1,161	86,082	206,318
Los Angeles	774	816	704	1,534	201,400	615,500
Miami	364	437	647	1,373	113,200	310,700
New York	6,545	6,848	720	1,536	239,862	685,002
Washington	307	395	837	1,700	181,933	477,031
C. Income, Education and Driving						
	Av. Household Income		Share, BA or more		Share Drive to Work	
	2000	2020	2000	2020	2000	2020
Boston	72,196	127,262	0.37	0.489	0.711	0.644
Chicago	67,437	103,463	0.29	0.39	0.704	0.683
Houston	61,050	99,268	0.265	0.336	0.77	0.789
Los Angeles	61,811	103,220	0.249	0.335	0.704	0.721
Miami	52,753	82,379	0.217	0.307	0.738	0.757
New York	58,505	107,000	0.274	0.391	0.249	0.223
Washington	80,642	139,109	0.425	0.52	0.675	0.632

Sources: Decennial Census, 2000 and American Community Survey, 5-year data, 2016-2020.
Notes: All data are at the county level. When necessary, we aggregate to the CoStar market level. Population density is people per square kilometer. Income, rent and house value are in nominal 1999 and 2019 dollars. Share White and share Black are based on the number of White and Black non-Hispanic people. In 2020, we use only those who indicate "Black Alone." Rent is "median gross rent." We use markets as defined by CoStar, which in most cases approximate a county or aggregates of counties.