ISSN 1936-5330

Research Working Papers

# A Better Delineation of U.S. Metropolitan Areas

Jordan Rappaport and McKenzie Humann

April 2025; updated May 2025
RWP 25-01
http://doi.org/10.18651/RWP2025-01

FEDERAL RESERVE BANK OF KANSAS CITY

# A Better Delineation of U.S. Metropolitan Areas*

Jordan Rappaport† McKenzie Humann‡

May 18, 2025

**Abstract**

Metropolitan areas are a fundamental unit of economic analysis. Broadly defined, they are unions of built-up locations near each other among which people travel between places of residence, employment, and consumption. Despite the importance of metropolitan areas, metropolitan Core-Based Statistical Areas and other official U.S. delineations considerably stray from this broad definition. We develop a simple algorithm to better match it, using commuting flows among U.S. census tracts in 2000. Three judgmental parameters govern the threshold strength of commuting ties between locations to include them in the same metropolitan area, the maximum separating distance between locations, and the threshold density of outlying settlement. A parameterization that balances encompassing commuting flows and excluding sparsely settled land delineates 361 Kernel-Based Metropolitan Areas (KBMAs), in aggregate capturing almost all the population and employment of metropolitan CBSAs in a small fraction of their land area. We benchmark KBMAs against two alternative parameterizations, one that prioritizes encompassing commuting flows and one that prioritizes excluding less built-up and less near locations.

**Keywords:** delineating metropolitan areas, commuting flows, city size, Core-Based Statistical Areas

**JEL Classification Numbers:** R12, R14, R23

*Thank you to Chris Cunningham, Giles Duranton, Xavier Gabaix, Todd Gardner, Sanghoon Lee, Albert Saiz, Elisabet Viladecans-Marsal, and several anonymous referees for helpful feedback. Thank you to Caleb Bray, Francis Dillon, and Sai Avinash Sattiraju for excellent research assistance. The views expressed herein are solely those of the authors and do not necessarily reflect the views of the Federal Reserve Bank of Kansas City or the Federal Reserve System.

†Corresponding author. Federal Reserve Bank of Kansas City. Email: jordan.rappaport@kc.frb.org

‡Independent Researcher Email: mckenzie.humann@gmail.com

# 1 Introduction

Metropolitan areas are a fundamental unit of economic analysis. We define them broadly as unions of built-up locations near each other with combined population of at least moderate scale and among which a significant share of residents and workers travel on a day-to-day basis between places of residence, places of employment, and places of consumption. By this definition, metropolitan areas correspond to distinct markets for labor and for non-traded goods and services. They are also likely to correspond to the geographic area determining many agglomerative externalities and for sharing many production and consumption amenities. In addition, national populations sort themselves across and within metropolitan areas with respect to numerous characteristics. Conversely, metropolitan areas may foster a sense of shared identity among diverse residents.

Despite this fundamental importance, official delineations considerably stray from our broad definition of metropolitan areas. Metropolitan Core-Based Statistical Areas (CBSAs), delineated by the U.S. Office of Management and Budget, vastly overbound metropolitan land area. For example, the Phoenix CBSA spans more than 14,000 square miles, mostly empty desert, and the Honolulu CBSA includes a Pacific atoll more than 900 miles from its downtown. Even so, many commonly recognized labor markets are split into multiple metropolitan CBSAs, including Raleigh and Durham, Los Angeles and Riverside–San Bernadino, and San Francisco and San Jose. Conversely, other arguably separate metropolitan areas are encompassed by a single metropolitan CBSA. For example, 20 miles of empty desert in 2000 separated the settled portion of the Coachella Valley—Palm Springs and neighboring municipalities in the central portion of the Riverside–San Bernadino CBSA—from the settled western portion of the CBSA.

Two other official U.S. delineations that are used as proxies for metropolitan areas egregiously stray from our definition. Urbanized Areas (UAs), combinations of census blocks with contiguous population density above a specified threshold, fragment unambiguous metropolitan areas. For example, a slight gap in residential settlement causes a portion of one of Kansas City's main suburban municipalities to be delineated as its own UA, notwithstanding that more than half of its employed residents work in the Kansas

City UA and more than one third of its employment is made up of workers living in the Kansas City UA. Commuting Zones (CZs), combinations of U.S. counties delineated to identify rural labor markets, both carve up some unambiguous metropolitan areas and arguably span multiple others. For example, six commuting zones in the vicinity of New York City are connected to Manhattan by commuter rail. Conversely, a commuting zone in Pennsylvania fully encompasses four metropolitan CBSAs (Harrisburg-Carlisle, Lancaster, Lebanon, and York-Hanover).

The failures of official delineations impede decision making, policy implementation, and fundamental understanding. Most immediately, metropolitan delineations affect a wide range of choices, including on infrastructure investment, transportation planning, corporate location, and government-sponsored development. In addition, implementing a slew of legislation depends closely on metropolitan delineations (Congressional Research Service, 2014). For example, the income ceiling in 2024 for a neighborhood to qualify as low income under the Community Reinvestment Act was 13 percent lower in the Durham CBSA than in the Raleigh CBSA (Federal Financial Institutions Examination Council, 2024).

From a research perspective, Duranton (2021) argues that "meaningful and appropriate" delineations of metropolitan areas are required "to understand anything about fundamental urban questions." Better metropolitan delineations are also likely to improve our understanding of many economic processes for which variation across locations underpins empirical study. Examples include labor supply and demand (e.g., Kennan and Walker, 2011; Autor, Dorn and Hanson, 2013; Monte, Redding and Rossi-Hansberg, 2018), housing supply and demand (e.g., Green, Malpezzi and Mayo, 2005; Saiz, 2010; Van Nieuwerburgh and Weill, 2010, Landvoigt, Piazzesi and Schneider, 2015), non-housing consumption demand (Mian and Sufi, 2012, 2014; Guren et al., 2021), productivity spillovers (Glaeser and Mare, 2001; Moretti, 2004; Combes, Duranton and Gobillon, 2008; Greenstone, Hornbeck and Moretti, 2010), and social mobility (Chetty et al., 2014).

To allow researchers, decision-makers, and others better address these and other questions, we develop a simple algorithm that delineates metropolitan areas that are consistent with our broad definition. Similar to the construction of CBSAs, UAs serve as anchoring cores. The algorithm iteratively joins cores together to form kernels and then builds out from the kernels to encompass surrounding census tracts. Two parameters set the threshold strength of pairwise commuting flows and the maximum separating distance to include locations in the same metropolitan area; a third parameter sets a threshold density for outlying settlement.

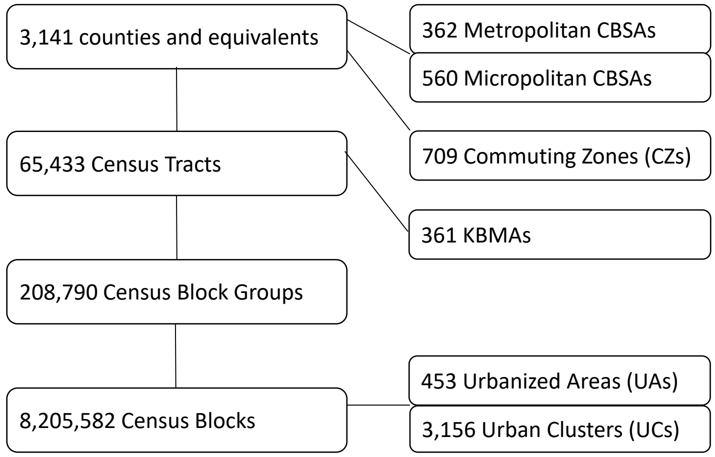
Our baseline parameterization balances encompassing commuting flows and excluding locations that arguably are not sufficiently near or built up. The resulting Kernel-Based Metropolitan Areas (KBMAs) are likely to be appropriate for most purposes, and so we hope they will be widely adopted. As alternative benchmarks, we delineate more expansive kernel-based metropolitan regions and more compact kernel-based urban areas: The parameterization for the metropolitan regions more heavily weights encompassing commuting flows and so may better match studying extended connections such as long-distance commuting and regional supply chains. The parameterization for the urban areas more heavily weights excluding less near and less built-up locations and so may better match studying narrow geographic spillovers and urban land usage. In addition, replication code allows users to parameterize delineations as they judge appropriate.

The 361 KBMAs we delineate encompass 88 percent of the aggregate population of metropolitan CBSAs and 94 percent of their aggregate employment in just 15 percent of their aggregate land area. Their population ranges from 50,000 to 19 million and their land area from 27 to 6,100 square miles. The population distributions of the kernel-based metropolitan regions and kernel-based urban areas closely match that of KBMAs. Under all three parameterizations, land area expands less than proportionately with population, consistent with centripetal forces, such as centralized employment and amenities, constraining metropolitan expansion.

# 2 Existing Metropolitan Delineations

Delineating metropolitan areas requires making a number of judgments, either explicit or implicit. What geographic units serve as a building blocks? What makes building blocks sufficiently integrated to join them together in a cluster? Should clusters be

**Figure 1: Metropolitan Building Blocks Using the 2000 Decennial Census**




Note: excludes U.S. territories.

anchored by a core? What qualifies a cluster as metropolitan?

## 2.1 Building Blocks

Census blocks are the most granular geographic unit for which the Census Bureau collects data. Blocks are typically quite small, for example the rectangle bounded by one city block on each side, but can also span many square miles in sparsely settled areas. For the 2000 decennial census, the Census Bureau delineated approximately 8.2 million blocks with land area ranging from 0 (all water) to 8,072 square miles (unsettled), with a median of 0.01 square miles. Their population ranged from 0 (more than a third of blocks) to a bit over 23,000. Across census blocks with residents, the median population was 25. Only limited block data is publicly available.

Census block groups combine up to 1,000 blocks and serve mainly for statistical purposes. Specifically, they are the smallest geographic unit for which the Census Bureau reports tabulated data from decennial census sample questions and from the American Community Survey.

Census tracts combine census block groups with several goals, including maintaining relatively stable boundaries over time (U.S. Census Bureau, 1997*b*). Even so, numerous tracts are re-delineated in preparation for each decennial census. For the 2000 census, the

Census Bureau targeted tracts to encompass between 1,500 and 8,000 residents. Nevertheless, 2 percent had population in 2000 below 500 and a handful had population above 20,000. Tracts' land area in 2000 ranged from 0.06 square miles at the 1st percentile to 800 at the 99th percentile (median = 1.96 square miles). The endogenous delineation of census tracts to meet the population targets induces a tight negative correlation between land area and population density.

Counties in 2000 combined between 1 and 2,100 tracts, with population ranging from 1,000 at the 1st percentile to 1.1 million at the 99th percentile (median = 25,000) and land area ranging from 10 square miles at the 1st percentile to 8,100 square miles at the 99th percentile (median = 616). Importantly, their average land area considerably differs across U.S. regions, ranging from a median of 425 square miles in the South Atlantic census division to a median of 2,275 in the Mountain census division; this variation suggests that metropolitan delineations constructed using counties as building blocks are not comparable across different parts of the U.S. On the other hand, county borders have remained relatively stable since 1920, making counties an ideal building block for delineating metropolitan areas with unchanged borders since then. In addition, a wealth of county data is available from government agencies and private organizations.

## 2.2 Urbanized Areas and Urban Clusters

Urbanized Areas were first delineated by the Census Bureau following the 1950 decennial census, partly in response to rapidly growing population just outside the boundaries of large incorporated municipalities (Ratcliffe, 2015). Beginning with the 2000 decennial census, the Census Bureau used a mostly granular algorithm to delineate UAs and a smaller counterpart, Urban Clusters (UCs), as combinations of census blocks (U.S. Census Bureau, 2002*b*,*a*).[1]

The granular portion of the UA/UC algorithm focuses on residential density and proximity. It starts by identifying clusters of two or more contiguous block groups or blocks with population density of at least 1,000 per square mile. Expanding outward, adjacent

[1]The separate designations were dropped following the 2020 decennial census, in favor of "Urban Area", and the threshold population for inclusion was increased from 2,500 to 5,000 (U.S. Census Bureau, 2022).

block groups and blocks with population density of at least 500 are attached. Each resulting cluster is then joined with any similarly constructed clusters separated from it by no more than 0.5 miles along a connecting road segment, forming *interim cores* of chained clusters. Next, interim cores separated by no more than 2.5 miles along a connecting road segment are joined together, forming chained cores. Additional criteria add in blocks mostly surrounded by the chained cores as well as adjacent blocks containing a major airport. The resulting unions of census blocks constitute *urban area agglomerations*.[2]

All urban area agglomerations with population between 2,500 and 50,000 qualify as UCs under the 2000 algorithm, and most with population above this qualify as UAs. However, some of the larger agglomerations are split into two or more UAs to backwardly align them with metropolitan statistical areas (MSAs) delineated during the 1990s. For example, the Los Angeles UA is contiguous with two other UAs, Riverside–San Bernadino and Thousand Oaks, and so the three constitute a single urban area agglomeration. Splitting it aligns each of the resulting UAs with a legacy MSA. This backward alignment to the 1990s is perpetuated in the algorithms constructing UAs following the 2010 and 2020 decennial census, which split urban area agglomerations to avoid combining UAs previously delineated as separate (U.S. Census Bureau, 2011, 2022).[3]

Backwardly aligning the UAs in 2000 with legacy MSAs breaks an otherwise granular algorithm, contaminating using UAs for some purposes. In addition to the explicit history dependence, doing so introduces ad hoc subjectiveness that was applied idiosyncratically across locations. Specifically, metropolitan delineations during the 1990s relied heavily on local opinion—the views of a wide range of public groups including business and other leaders, chambers of commerce, planning commissions, and local officials—to partition larger MSAs, designated as *consolidated*, into two or more *primary* MSAs (Office of Management and Budget, 1990).

[2]The term "urban area agglomeration" was introduced following the 2020 decennial census (U.S. Census Bureau, 2022).

[3]Backwardly aligning UAs has required more extensive splits of urban area agglomerations as urban settlement has spread. For UAs delineated following the 2000 decennial census, we identify 32 pairs with contiguous borders, ranging from a single point to a maximum of 9 miles, implying that splits were made at each. Eight of these contiguous borders exceeded 3 miles in length, the maximum prescribed by the published rule for the 2000 delineations (U.S. Census Bureau, 2002*b*). For the UAs delineated using the 2020 decennial census, we identify 121 pairs with contiguous borders, ranging in length up to 47 miles.

From a pragmatic perspective, a key flaw of UAs and UCs is that they fragment many locations that are arguably integrated. For example, as illustrated in Figure 2, five separate UAs are located along a 25-mile ray in the southern portion of the Oxnard–Thousand Oaks CBSA. Large commuting flows among three of these—Los Angeles, Simi Valley, and Thousand Oaks—far exceed our threshold metric for combining them in a single metropolitan area, as do the commuting flows between the Oxnard and Camarillo UAs. Similarly, as described in the introduction, a slight gap in residential settlement causes a portion of one of Kansas City's main suburban municipalities to be delineated as its own UA, notwithstanding that more than half of its residents work in the Kansas City UA and more than one third of its employment is made up of workers living in the Kansas City UA. Analogous suburban UAs, from which at least half of employed residents commute to a much larger UA, were delineated adjacent to the Phoenix, Seattle, Chicago, and New York UAs.[4]

Conversely, the relatively low density threshold for inclusion in urban area agglomerations contributes to some chains that arguably span multiple metropolitan areas. For example, a single urban area agglomeration stretches from northern New Jersey up along the Atlantic coast almost to New London, including a branch extending from New Haven north to Springfield, MA.

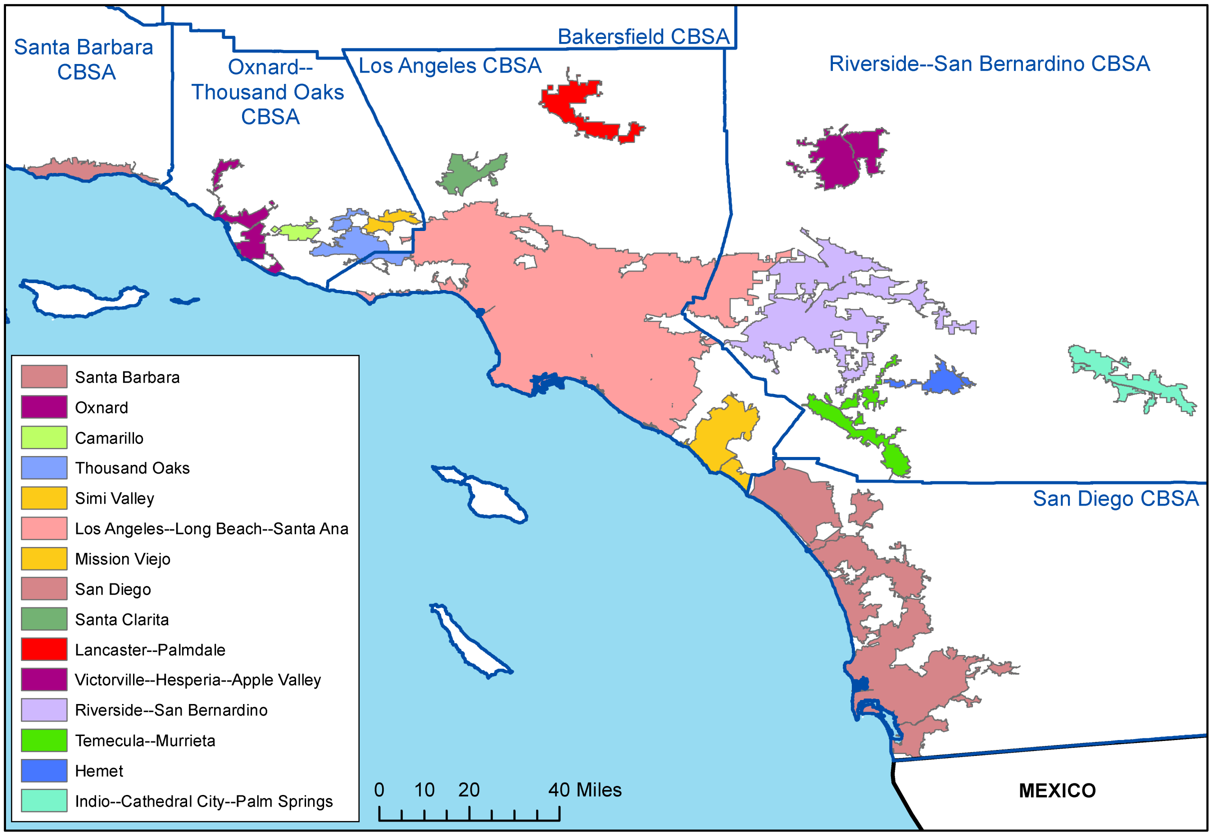
## 2.3 Metropolitan Core-Based Statistical Areas

The predecessor to the Office of Management and Budget (OMB) began delineating metropolitan statistical areas in the late 1940s, motivated by the desire to have geographic units for which government agencies could collect, tabulate, and publish data. The specific criteria and name of the delineated units evolved over time (Berry, Goheen and Goldstein, 1969; Office of Management and Budget, 1998; Gardner, 1999, 2021).

In preparation for the 2000 decennial census, OMB significantly revised its criteria

[4]The Lee's Summit, MO UA is separated from the Kansas City UA by 0.4 miles (implicitly not along a connecting road segment as otherwise they would not be separate). The Avondale, AZ UA is separated from the Phoenix–Mesa UA by 0.5 miles; the Marysville, WA, UA is separated from the Seattle UA by 0.4 miles; the Round Lake Beach–McHenrey–Grayslake, IL-WI UA is separated from the Chicago UA by 0.4 miles; and the Hightstown, NJ UA is separated from the New York–Newark UA by 0.1 miles.

Figure 2: Urbanized Areas in Southern California




Notes: Borders represent delineations following the 2000 decennial census. Blue lines demarcate metropolitan CBSAs. The Los Angeles UA includes a narrow corridor of census blocks along its northwest coast, some of which are not visible.

to delineate metropolitan and micropolitan Core-Based Statistical Areas (Office of Management and Budget, 2000). Both are constructed using counties as geographic building blocks, with UAs serving as cores anchoring metropolitan CBSAs and UCs with population of at least 10,000 serving as cores anchoring micropolitan CBSAs. Each of these UAs and UCs is associated with one or more *central counties*, in which they are substantially encompassed. Nearby counties are designated as *outlying counties* of the CBSA if at least 25 percent of their employed residents work in the central counties or if at least 25 percent of their employment is made up of workers who live in the central counties. The published criteria also allow OMB to combine adjacent groups of counties that are algorithmically delineated as separate CBSAs if local opinion favors doing so.

As summarized in Table 1, OMB delineated 362 metropolitan CBSAs in 2000 with

**Table 1: Official Delineations of U.S. Metropolitan Areas**

| Delineation | Core | Building Blocks | Criteria | obs | Population min | Population mdn | Population max | Land Area (sq.mi) min | Land Area (sq.mi) mdn | Land Area (sq.mi) max |
|---|---|---|---|---|---|---|---|---|---|---|
| Urbanized Areas (UAs) | none | census blocks | proximity, pop density | 453 | 50.1k | 117k | 17.8m | 12.1 | 64 | 3.4k |
| Urban Clusters (UCs) | none | census blocks | proximity, pop density | 3,156 | 2.5k | 5.9k | 49.4k | 0.04 | 4.1 | 273 |
| Core-Based Statistical Areas (CBSAs): | | | | | | | | | | |
| metropolitan | UA | counties | commuting, proximity | 362 | 52.5k | 223k | 18.3m | 140 | 1.6k | 27.3k |
| micropolitan | UC w pop $\geq$10k | counties | commuting, proximity | 560 | 13.0k | 43.7k | 182k | 110 | 770 | 21.4k |
| Commuting Zones (CZs) | none | counties | commuting | 709 | 13.9k | 175k | 16.4m | 620 | 3.6k | 166k |

Notes: Delineations are based on the 2000 decennial census and exclude U.S. territories. The CBSA delineations are the version promulgated on June 6, 2003 (Office of Management and Budget, 2003); the Commuting Zone delineations are disseminated by the Economic Research Service (Economic Research Service, 2012).

population ranging from 52,500 to 18.3 million and 560 micropolitan CBSAs with population ranging from 13,000 to 182,000. The considerable overlap in population between the two types emphasizes that they are distinguished from each other based on the population of their cores. OMB retained essentially the same criteria for the 2010 and 2020 decennial censuses (Office of Management and Budget, 2010, 2021). It also periodically updates delineations based on intercensal population estimates.

Metropolitan CBSAs deviate from our metropolitan definition in several ways. First, many vastly overbound metropolitan land area, reflecting the use of counties as building blocks. As described in the introduction, the vast majority of the Phoenix CBSA is empty desert. Similarly, the Riverside–San Bernadino and Anchorage CBSAs each span land area more than three times the size of Massachusetts. Even in the New York City–Newark–Edison CBSA, half of the land area in 2000 was accounted for by census tracts with both population and employment density below 500, the threshold for a census block to be included in a UA or UC.

Second, some CBSAs fully encompass what are arguably separate metropolitan areas. For example, the settled portion of the Coachella Valley constitutes its own television broadcasting market, contributing to our judgment that it belongs to a different metropolitan area than the Riverside–San Bernadino UA, which is a part of the Los Angeles television market.

Third, many CBSAs underbound metropolitan areas. For example, the rule splitting urban area agglomerations into multiple UAs causes Los Angeles and Riverside–San Bernadino to be delineated as separate CBSAs as it also does for San Francisco and San Jose. For Raleigh and Durham NC, a slight gap in residential settlement between them in 2000 causes them to be delineated as separate UAs, a necessary condition for them to be included in separate CBSAs. Especially egregious, 39 percent of workers residing in the Hinesville–Fort Stewart, GA metropolitan CBSA in 2000 commuted to places of employment in the Savannah, GA metropolitan CBSA.[5]

For research purposes, the CBSA delineations suffer from several methodological flaws. One is that they are history dependent. As described previously, metropolitan CBSA cores, UAs, are split to backwardly align with MSAs delineated during the 1990s. This history dependence deepened with the 2010 and 2020 delineations, reflecting that suburban expansion filled in unsettled gaps between UAs, thereafter requiring more extensive splits. Worse, the backward alignment is to delineations that relied heavily on local opinion, contaminating the CBSA algorithm.[6] Another flaw arises from the considerable variation in counties' average land area across U.S. regions, described previously, which implicitly delineates CBSAs differently across regions and so limits comparisons.

[5]The online supplemental materials include a table enumerating more than 200 pairs of metropolitan CBSAs with at least moderate cross commuting. The high commuting outflow rate from Hinesville–Fort Stewart to Savannah does not meet the 25 percent threshold for merging because a significant portion of it originates in an outlying rather than central county. In recognition of overbounding and underbounding, OMB also delineates subsets of CBSAs, Metropolitan Divisions, and supersets of them, Combined Statistical Areas. It also constructs analogs for the six New England states using county subdivisions as geographic building blocks.

[6]Illustrating this heavy dependence on local opinion, the 1990s algorithm delineated a Consolidated MSA chaining from southern New Jersey up through New York City and then further north into Connecticut and eastern Pennsylvania. OMB relied on local opinion to divide it into 15 Primary MSAs. Among the more idiosyncratic, the Jersey City Primary MSA is made up of a 14-mile strip of land tightly wedged between Newark, NJ and the Hudson River.

## 2.4 Additional Methodologies

A third set of possible proxies for U.S. metropolitan areas, Commuting Zones (CZs), were delineated by the U.S. Department of Agriculture's Economic Research Service following each of the 1980, 1990, and 2000 decennial censuses with the goal of better understanding rural labor markets (Tolbert and Sizer, 1987, 1996; Foote, Kutzbach and Vilhuber, 2017). Like CBSAs, CZs are constructed using counties as geographic building blocks. The commuting strength between two locations is calculated as the sum of the commuting flows in both directions normalized by the level of employment in the location where it is smaller. The delineation algorithm begins by constructing an N-by-N symmetric matrix of commuting strength between all possible pairs of counties; the pair with the strongest tie are joined, constituting a first cluster. The process is iteratively repeated, calculating a new (N-1)-by-(N-1) symmetric matrix and joining the pair of observations with the strongest tie until it falls below a judgmental threshold. The resulting clusters fully partition the United States.

CZs share many of the same flaws as CBSAs, including over-bounding and under-bounding. A large share of CZs encompass one, two, or three CBSAs together with a handful of adjacent rural counties. Even so, many medium and large CBSAs are split across multiple CZs, with groups of unambiguous suburbs separated from central business districts. As illustrated in the appendix, six CZs in the vicinity of New York City are connected to Manhattan by commuter rail.

Expanding beyond the United States, the European Union and the OECD delineate Functional Urban Areas in their member countries (Dijkstra and Poelman, 2012; Brezzi et al., 2012; Dijkstra, Poelman and Veneri, 2019). The first of four construction stages joins adjacent grid cells of 1 square kilometer into the same cluster if each has population density of at least 1,500 persons per square kilometer (3,885 persons per square mile). Resulting clusters that have population of at least 50,000 are considered *urban centres.* These are analogous to U.S. Urbanized Areas except that they have a density threshold almost eight-fold higher and so exclude considerable suburban area. The second stage combines the governmental administrative units that overlap each urban center, forming *core cities*. The third stage joins core cities between which commuting ties exceed a threshold strength, in essence creating kernels. The final stage builds out from the kernels, attaching local government administrative units with which commuting ties exceed a threshold strength.[7]

Duranton (2015) goes a step beyond the examples above, granularly delineating metropolitan areas that are multilevel hierarchies. The directional strength between locations is measured by the commuting outflow rate, calculated as the flow from origin to destination relative to the number of workers residing in the origin. A first round of joins classifies each location as subsidiary to the one with which it has the highest outflow rate, contingent on the rate exceeding a judgmental threshold. The resulting first-round clusters mix hub-and-spoke (B and C each attached to A) and chained configurations (C attached to B and B attached to A). These are then iteratively used to effect a second round of joins, attaching each first-round cluster to the one with which it has the strongest outflow rate. The iteration continues until no outflow rate exceeds the threshold. Applying the algorithm to U.S. counties constructs hierarchical clusters whose population is tightly correlated with that of matched CBSAs (Dingel, Miscio and Davis, 2021). Although CBSAs poorly match our broad definition of metropolitan areasa, UAs and UCs could be used as building blocks to construct a hierarchical version of our metropolitan kernels.

Reliable commuting data is not available for most nations, and so delineating metropolitan areas must rely on other approaches. One uses satellite images of nighttime lights and land cover. For example, Ch, Martin and Vargas (2021) identify clusters of adjacent pixels with nighttime light intensity above alternative thresholds to delineate between 4,200 and 6,700 metropolitan areas worldwide. Dingel, Miscio and Davis (2021) similarly use nighttime light to cluster pixels, which they match with intersecting admin-

[7]The European Union, OECD, the United Nations, and The World Bank jointly adopted an urban-rural classification system that complements the EU/OECD delineations (Eurostat et al., 2021; Dijkstra et al., 2021). Countries are gridded into cells of 1 square kilometer, each of which is assigned to one of three categories: an *urban centre*, constructed as described; an *urban cluster*, contiguous grid cells, not in urban centres, that have population density of at least 300 persons per square kilometer (777 persons per square mile) and total population of at least 5,000; and *rural grid cells*, all cells not in an urban centre or urban cluster. These are respectively characterized as "densely populated", "intermediate density", and "thinly populated".

istrative units. Using U.S. counties as administrative building blocks, the implied unions have population that is tightly correlated with the population of matched CBSAs across a range of light-intensity thresholds, validating using the methodology for some purposes when commuting data is not available.

Another approach focuses on physical structures. de Bellefon et al. (2021) classify small grid cells in France as urban if the density of buildings in them exceeds a threshold percentile across all grid cells in a country and then cluster the grid cells to construct urban areas. Similarly, Arribas-Bel, Garcia-Lopez and Viladecans-Marsal (2021) apply a spatial clustering algorithm to the precise location of all buildings in Spain to construct urban areas.

A third approach relies more directly on judgment. Galdo, Li and Rama (2021) ask several groups of assessors to use Google Earth and Google Maps to classify a sample of locations in India as either urban or rural. A machine learning algorithm then uses several measurable characteristics of the locations—such as population, population density, land cover, and nighttime luminosity—to predict the urban/rural status of more than 500,000 villages, towns, and cities.

# 3 Constructing KBMAs

Our metropolitan definition is purposely imprecise, leaving scope for judgment on whether specific delineations are consistent with it. Alternative parameterizations of the KBMA algorithm can match a wide range of judgments on encompassing places of residence and employment that are near each other while excluding lightly settled locations. Like the methodologies described above, no explicit component is included to encompass places of consumption.

## 3.1 Algorithm

We use census tracts, UAs, and UCs as our building blocks, taking advantage of carefully measured commuting flows among them. As with CBSAs, UAs and UCs with population above 10,000 serve as the cores of KBMAs. The first stage of the algorithm iteratively joins these cores together into kernels, paralleling the iterative construction of CZs. Doing so combines many of the UAs that are split by backward alignments. The second stage builds out from the kernels, attaching outlying tracts tied to them by sufficiently strong commuting flows.

We symmetrically measure the strength of the commuting tie between two locations by the sum of the inflow and outflow rates in both directions. Let $f_{i,j}$ represent the gross commuting flow from location $i$ to location $j$; let $e_i$ and $w_i$ respectively represent employment in location $i$ and workers residing there. The commuting strength between $i$ and $j$ (and between $j$ and $i$) is given by,

$$s_{i,j}, s_{j,i} = \frac{f_{i,j}}{w_i} + \frac{f_{i,j}}{e_j} + \frac{f_{j,i}}{w_j} + \frac{f_{j,i}}{e_i} \tag{1}$$

Measuring strength symmetrically abstains from imposing a hierarchy. Normalizing directional flows both by resident workers and employment equally weights the importance of a location serving as a source of labor supply and as a source of labor demand.

Pragmatically, summing the four terms rather than taking the maximum (the CBSA formula) avoids a bias against combining locations of similar size. The normalization of the gross flows yields relatively low values for all four terms when two locations have similar size compared to the maximum value of the four terms when one location is much larger than the other. Correspondingly, we judge that appropriately delineating metropolitan areas when measuring strength by the maximum value would require setting a much lower threshold, $\sigma$, for joining two similar locations than for joining two unequal ones.

The initial iteration of the first stage begins by constructing an N-by-N symmetric matrix of the commuting strength between all pairs of cores. The pair with maximum strength, subject to having separating distance no more than $\delta$, is joined together and then a new (N-1)-by-(N-1) matrix is constructed. Iterating continues until the maximum pairwise strength drops below $\sigma$.

The second stage attaches outlying tracts to a kernel if the strength of their commuting tie with the kernel weakly exceeds $\sigma$, their distance from the kernel does not exceed $\delta$, and either their population or employment density weakly exceeds $\eta$.

We rely almost exclusively on one source of data, the Census Transportation Planning Package 2000 (Bureau of Transportation Statistics, 2005). It re-tabulates responses to the journey-to-work questions on the long form of the 2000 decennial census by tract of employment and by origin-destination pairs.[8]

As a prerequisite to implementing our algorithm, we construct tract-based approximations of UAs and UCs with population above 10,000. Most census tracts are either fully overlapped by a UA/UC or else fully disjoint with any UA/UC; even so, many tracts partly intersect with a UA/UC. As described in the appendix, we use a fitness criterion to set threshold shares of a tract's population and land area that must intersect with a UA/UC to include the tract in that UA/UC's approximation. Only 1,271 of the 1,372 UAs and eligible UCs are actually approximated, reflecting that 101 of the eligible UCs have no intersecting tract that meets both the population and land-area thresholds. As described in the appendix, this shortfall only negligibly affects delineations.

## 3.2 Parameterization

Parameterizing the threshold commuting strength and maximum allowed separating distance to join locations and the threshold density of outlying tracts to attach them to kernels rests heavily on judgment. Any set of values will inevitably contradict many people's priors, either because of specific implied delineations or because of judgements of what qualifies as sufficiently integrated, near, and built-up. Alternative parameterizations illuminate how judgments affect delineations.

We judgmentally parameterize KBMAs to balance encompassing commuting flows and excluding sparsely settled land, while resting only lightly on separating distance to constrain joins. As such, KBMAs serve as a baseline likely to match most reasons for using metropolitan delineations. We also delineate more expansive kernel-based metropolitan regions, setting lower values for threshold commuting strength and density and a higher

[8] A question on the 2000 decennial census long form and subsequent American Community Surveys asks respondents to give the address where they primarily worked the previous week. An alternative source of origin-destination commuting flows, the Census Bureau's LEHD Origin-Destination Employment Statistics (LODES), determines workplaces by the address of the business establishment with which employees are associated for payroll reporting. These establishment locations frequently differ from actual work locations. In addition, the LODES dataset does not cover non-payroll employment.

allowed separating distance, and more compact kernel-based urban areas, setting higher values for threshold strength and density and a lower allowed separating distance.

The kernel portion of the construction depends only on the commuting strength and distance parameters, $\sigma$ and $\delta$. The composition of many kernels is especially sensitive to the parametrization of $\sigma$, which we set to 0.25 for KBMAs based on our judgment of the implied unions. To span a wide range of judgments, we set $\sigma$ much lower, to 0.10, for kernel-based metropolitan regions and much higher, to 0.40, for kernel-based urban areas. We set the maximum allowed separating distance, $\delta$, to 20 miles for KBMAs, measured between the nearest tract centroids, consistent with our judgment of "near" and sufficiently high that it blocks relatively few joins during the kernel iteration.[9] Our alternative metropolitan region and urban area parameterizations respectively set $\delta$ to 40 miles and to 10 miles. Under the KBMA parameterization, iterating ends after 340 joins, leaving 931 kernels.

Prior to iterating, the 1,271 cores form more than 800,000 unordered pairs. Constrained to those separated by no more than 20 miles, the initial iteration joins the Miami and Key Biscayne cores, which have pairwise commuting strength of 1.28 and separating distance of 5.2 miles. Table 2 illustrates some of the later iterations as $\sigma$ is incrementally lowered from a strength of 0.30 to its KBMA parameterized value of 0.25 and then further down to a strength of 0.20. For example, the 0.30 strength threshold is sufficiently low to join Raleigh and Durham with one additional core; San Francisco and San Jose with eight additional cores; and Los Angeles, Riverside–San Bernadino, and Mission Viejo with 11 additional cores. Lowering $\sigma$ to 0.25 joins the already combined Manchester and Nashua cores with Boston, Worcester, and two other cores. Further lowering $\sigma$ to 0.20 joins this cluster with Barnstable Town (Cape Cod) and three other cores. Doing so also joins the Washington D.C. and Baltimore clusters, as well as the Los Angeles and Oxnard clusters. It would additionally join the Indio–Cathedral City–Palm Springs core to the Los Angeles cluster except that the separating distance between them is a few tenths above the allowed

[9]The tract centroids we use are the *internal points* reported by the Census Bureau. Some irregular-shaped tracts have geographic centroids that lie outside their physical territory, e.g. crescent shapes and tracts split by a water body. In these cases, the Census Bureau locates the internal point in the tract portion nearest the geographic centroid.

**Table 2: Kernels under Alternative Commuting Strength Thresholds**

| **Tighter** ($\sigma$=0.30) | **Actual** ($\sigma$=0.25) | **Looser** ($\sigma$=0.20) |
|---|---|---|
| Raleigh, NC; Durham, NC; 1 other | | Smithfield, NC |
| San Francisco-Oakland, CA; San Jose, CA; 8 others | Fairfield, CA; Vacaville, CA; Napa, CA; Dixon, CA | |
| Los Angeles-Long Beach-Santa Ana, CA; Riverside-San Bernadino, CA; Mission Viejo, CA; 11 others | | Oxnard, CA; Camarillo, CA; Santa Paula, CA |
| Oxnard, CA; Camarillo, CA; Santa Paula, CA | | *included in Los Angeles kernel* |
| Nashua, NH-MA; Manchester, NH | Boston, MA-NH-RI; Worcester, MA-CT; Leominster-Fitchburg, MA; Athol, MA | Barnstable Town, MA (Cape Cod); Dover-Rochester, NH–ME; Portsmouth, NH–ME; 2 others |
| Washington, DC-VA-MD; Frederick, MD; 4 others | | Baltimore, MD; Aberdeen-Havre de Grace-Bel Air, MD; 3 others |
| Baltimore, MD; Aberdeen-Havre de Grace-Bel Air, MD; 4 others | | *included in Washington DC, kernel* |
| Seattle, WA; Marysville, WA; North Bend, WA | | Bremerton, WA; Olympia-Lacey, WA; Centralia, WA; Shelton, WA |
| Detroit, MI; Ann Arbor, MI; South Lyon-Howell-Brighton, MI; Monroue, MI | Lapeer, MI | Port Huron, MI |
| New York-Newark, NY-NJ-CT; Trenton, NJ; Hightstown, NJ (Princeton); 8 others | Poughkeepsie-Newburgh, NY; Port Jervis, NY–PA; New Paltz, NY | |
| | Providence, RI-MA; New Bedford, MA | |
| | Akron, OH; Canton, OH; Alliance, OH | |
| | | Johnson City, TN; Kingsport, TN-VA |

Notes: The left column reports cores joined in selected kernels under a commuting strength threshold, $\sigma$, of 0.30, an increment above the KBMA parameterized value. The middle and right columns report additional cores joined from lowering the strength threshold to its KBMA value ($\sigma = 0.25$) and then lowering it a further increment. Joined cores must be separated by no more than 20 miles ($\delta = 20$).

20 miles.

We judge that setting $\sigma$ to a commuting strength of either 0.30 or 0.25 constructs a set of kernels that is consistent with our metropolitan definition. One reason we choose the lower value is to keep the threshold comfortably below the join of the Raleigh and Durham cores, which occurs at a strength slightly above 0.30. We are more skeptical that the set of kernels implied by lowering $\sigma$ to a strength of 0.20 is consistent with our definition. A number of the incremental joins are with cores we think of as seasonal rather than day-to-day destinations. Specifically, Oxnard, Cape Cod, Olympia, and Port Huron all market themselves as coastal getaways.

In addition, setting $\sigma$ to a commuting strength of 0.25 rather than 0.20 implies a more robust buffer against joining the Indio–Cathedral City–Palm Springs core to the Los Angeles kernel, consistent with our prior that it constitutes its own metropolitan area. Otherwise, not joining the two rests fragilely on their separating distance being a tick above the parameterized maximum of 20 miles. More generally, we methodologically prefer commuting strength to serve as the primary determinant of joins, reflecting that strength typically decreases with separating distance and so in part implicitly accounts for distance.

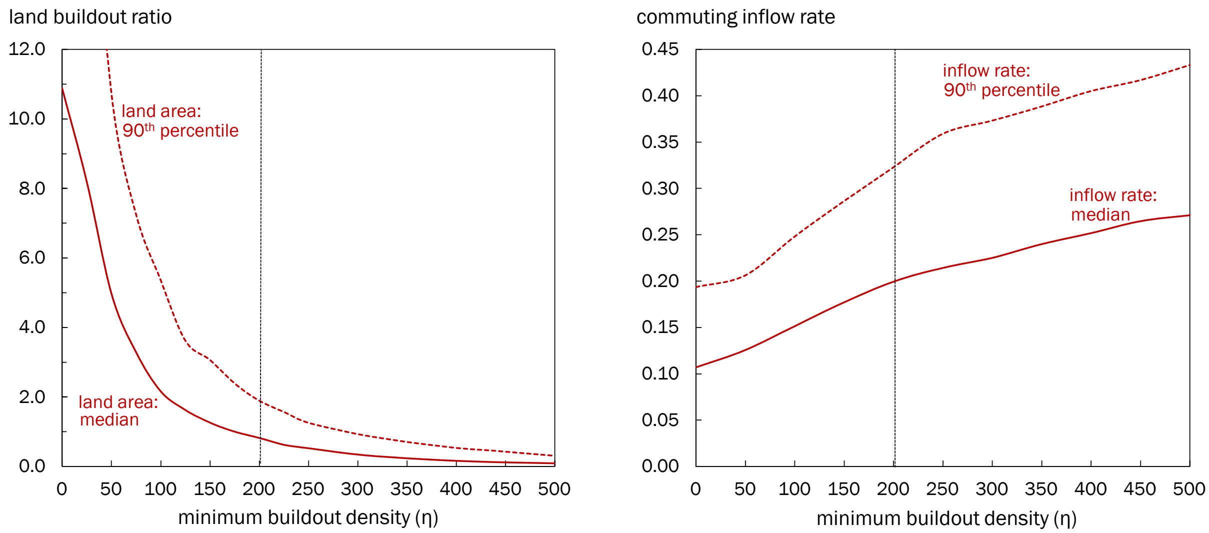
We judge that the kernel iterations appropriately recombine UAs split from each other by the backward alignment with the 1990s MSAs. We are able to identify 32 such splits of urban area agglomerations, 19 of which are recombined at the KBMA parameterization. Several other pairs are recombined if $\sigma$ is set to a strength of 0.20, including Port Huron and Detroit, Portsmouth and Boston, and Johnson City and Kingsport. All but four pairs are recombined at the parameterization for kernel-based metropolitan regions, which sets $\sigma$ to a strength of 0.10.

Not rejoining the remaining four pairs of split UAs illustrates the success of the kernel iteration at avoiding constructing long chains of cores. For example, one of the remaining splits divides New York City and Bridgeport–Stamford into separate UAs, between which the pairwise commuting strength is 0.19. The iterative sequencing initially joins each to other cores, diluting the pairwise strength between the resulting kernels. Further iterative dilutions delay combining New York City and Bridgeport–Stamford in the same kernel until $\sigma$ falls below a strength of 0.08.

In contrast to the high sensitivity for specific kernels, an appendix figure illustrates the gradual decline in the number of kernels as $\sigma$ is lowered. It also illustrates that relatively few iterative joins are constrained by the maximum allowed separating distance, $\delta$, at the parameterizations of KMBAs, kernel-based metropolitan regions, and kernel-based urban areas.

The buildout stage requires additionally parameterizing $\eta$, the minimum settlement density for eligible tracts to be attached to a kernel. We judgmentally set it to 200

**Figure 3: Sensitivity of Built-Out Kernels to the Density Threshold**




Notes: Left panel shows the median and 90th-percentile ratios of the land area of the buildout portion of built-out kernels relative to the land area of the kernel portion as $\eta$ is increased from 0 to 500. The dashed vertical line corresponds to the parameterized KBMA value, $\eta = 200$. The right panel shows the median and 90th percentile rates of commuting inflows. For comparability, kernels are restricted to the 302 with population of at least 50,000. A complementary appendix figure illustrates buildout ratios for population and employment, and the commuting outflow rate.

(residents per square mile or workers per square mile, whichever is higher), balancing encompassing commuting flows and excluding sparsely settled land.

Figure 3 illustrates the tradeoff. The left panel shows the ratio of land in the buildout portion of KBMAs to land in their kernel portion as $\eta$ is increased from 0 to 500. The median and 90th percentile buildout ratios at a density threshold of 200 are 0.8 and 1.9, respectively. Both begin to blow up as $\eta$ is lowered below this. The right panel shows the corresponding median and 90th percentile of KBMAs' commuting inflow rates. Both rise significantly as $\eta$ is increased from 0.

We judge that the marginal benefit of encompassing more flows by decreasing $\eta$ below 200 approximately equals the marginal cost of encompassing additional sparsely settled land area. To span a wide range of judgments, we drop the density criterion for kernel-based metropolitan regions, setting $\eta$ to 0. For the kernel-based urban areas, we set $\eta$ to 500, the threshold population density for census blocks to be included in a UA or UC. An analogous appendix figure shows population and employment buildout ratios and the encompassment of commuting outflows, all of which are less sensitive to the

**Table 3: 50 Largest KBMAs by Population in 2000**

| rank | KBMA Title | Population | rank | KBMA Title | Population |
|---|---|---|---|---|---|
| 1 | New York–Newark, NY–NJ | 19,139,000 | 26 | Orlando, FL | 1,657,000 |
| 2 | Los Angeles–Long Beach–Santa Ana, CA | 15,191,000 | 27 | Sacramento, CA | 1,614,000 |
| 3 | Chicago, IL–IN–WI | 8,831,000 | 28 | Kansas City, MO–KS | 1,498,000 |
| 4 | San Jose–San Francisco–Oakland, CA | 6,257,000 | 29 | Milwaukee, WI | 1,471,000 |
| 5 | Philadelphia, PA–NJ–DE–MD | 5,520,000 | 30 | Virginia Beach, VA | 1,443,000 |
| 6 | Boston, MA–NH–CT | 5,250,000 | 31 | San Antonio, TX | 1,414,000 |
| 7 | Miami, FL | 4,932,000 | 32 | Providence, RI–MA | 1,391,000 |
| 8 | Dallas–Fort Worth–Arlington, TX | 4,736,000 | 33 | Charlotte, NC–SC | 1,353,000 |
| 9 | Detroit, MI | 4,501,000 | 34 | Salt Lake City, UT | 1,339,000 |
| 10 | Washington, DC–VA–MD–WV | 4,428,000 | 35 | Las Vegas, NV | 1,322,000 |
| 11 | Houston, TX | 4,278,000 | 36 | Indianapolis, IN | 1,304,000 |
| 12 | Atlanta, GA | 3,962,000 | 37 | Columbus, OH | 1,295,000 |
| 13 | Phoenix–Mesa, AZ | 2,991,000 | 38 | New Orleans, LA | 1,220,000 |
| 14 | Seattle, WA | 2,942,000 | 39 | Buffalo, NY | 1,090,000 |
| 15 | San Diego, CA | 2,742,000 | 40 | Hartford, CT | 1,018,000 |
| 16 | Minneapolis–St. Paul, MN–WI | 2,640,000 | 41 | Austin, TX | 996,000 |
| 17 | Baltimore, MD–PA | 2,504,000 | 42 | Memphis, TN–MS–AR | 990,000 |
| 18 | Denver–Aurora, CO | 2,314,000 | 43 | Raleigh–Durham, NC | 978,000 |
| 19 | Tampa–St. Petersburg, FL | 2,313,000 | 44 | Nashville-Davidson, TN | 970,000 |
| 20 | St. Louis, MO–IL | 2,275,000 | 45 | Akron, OH | 928,000 |
| 21 | Cleveland, OH | 2,156,000 | 46 | Louisville, KY–IN | 917,000 |
| 22 | Pittsburgh, PA | 2,045,000 | 47 | Jacksonville, FL | 907,000 |
| 23 | Bridgeport–New Haven–Stamford, CT–NY | 1,844,000 | 48 | Oklahoma City, OK | 871,000 |
| 24 | Cincinnati, OH–KY–IN | 1,744,000 | 49 | Honolulu, HI | 864,000 |
| 25 | Portland, OR–WA | 1,729,000 | 50 | Richmond, VA | 854,000 |

Note: An enumeration of all KBMAs, with additional variables, is included in the paper's supplemental materials.

parameterization of $\eta$.

Lastly, we classify only the 361 built-out kernels that have population above 50,000 as KBMAs, matching the convention used by several of the delineations described in the previous section. Of course, higher or lower population thresholds may better match various purposes. Table 3 lists the 50 KBMAs with largest population.[10] An enumeration of all KBMAs along with detailed data on underlying census tracts, cores, built-out kernels, and the iterative joins constructing kernels are included in the paper's supplemental materials.

[10] We title KBMAs using the Census Bureau's algorithm for titling UAs and UCs (U.S. Census Bureau, 2002*b*). The first listed name is the incorporated place with highest population. The names of the next two incorporated places by population are included if they have population exceeding 250,000 or if their population is at least two thirds that of the largest incorporated place. State postal abbreviations are ordered to correspond to any places included in the title and then by descending order of each state's population in the KBMA.

# 4 Realized Delineations

To gauge our success in matching our metropolitan definition, we first show maps of realized KBMAs and describe KBMAs' overlap with metropolitan CBSAs. We then document the extent to which KBMAs encompass commuting inflows and outflows.

## 4.1 Illustrations

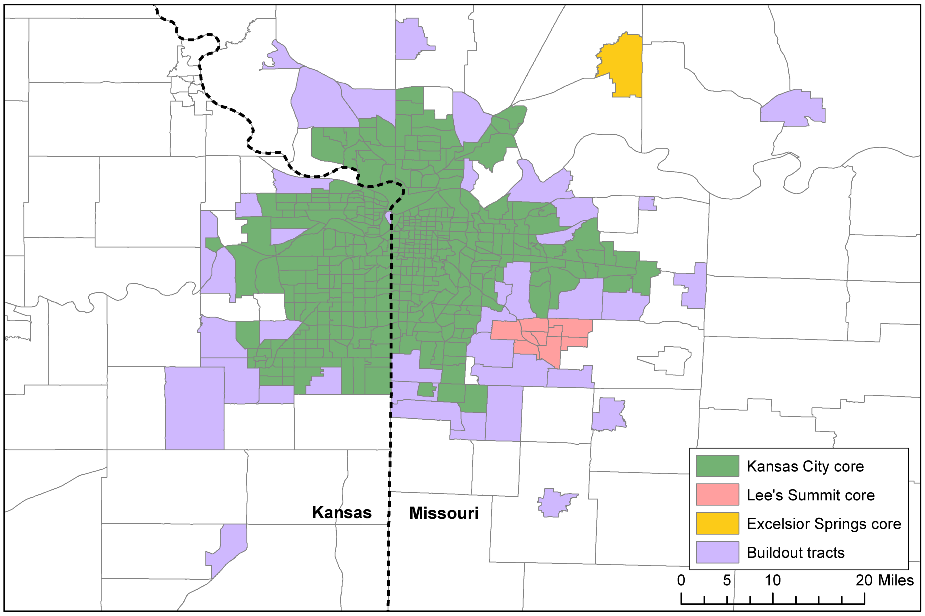
Figure 4 illustrates the composition of the Kansas City MO–KS KBMA. The kernel's three cores correspond to the Kansas City, MO–KS UA; a portion of a major suburb, Lee's Summit; and a small suburban UC, Excelsior Springs. Most of the buildout tracts are contiguously connected to the kernel; the detached ones correspond to UCs with population below 10,000. Most excluded tracts are expansive, reflecting sparse settlement. The main exception, the cluster of small tracts near the northwest corner, corresponds to the Leavenworth UC, whose commuting strength with the Kansas City kernel falls just short of the threshold to combine with it and whose built-out population on its own falls just short of the threshold to qualify as a KBMA.

Figure 5 zooms out. Kansas City and its three neighboring KBMAs each occupy the most densely settled portion of a corresponding metropolitan CBSA. The commuting strength between the Kansas City and Lawrence kernels is a tick below 0.150, sufficient to combine them under the parameterization for kernel-based metropolitan regions.

The correspondence between KBMAs and metropolitan CBSAs is less clean in Southern California. As illustrated in Figure 6, the Los Angeles KBMA encompasses essentially the entire population of the Los Angeles CBSA, most of the population of the Riverside-San Bernadino CBSA, and a significant portion of the population of the Oxnard-Thousand Oaks CBSA. The parameterization for kernel-based metropolitan regions, illustrated in the paper's supplemental materials, joins the Los Angeles, Oxnard, and Indio KBMAs together. This combination closely corresponds to an implicitly crowd-sourced judgement of "Greater Los Angeles", defined by Wikipedia as the combination of the Los Angeles, Riverside-San Bernadino, and Oxnard-Thousand Oaks CBSAs.

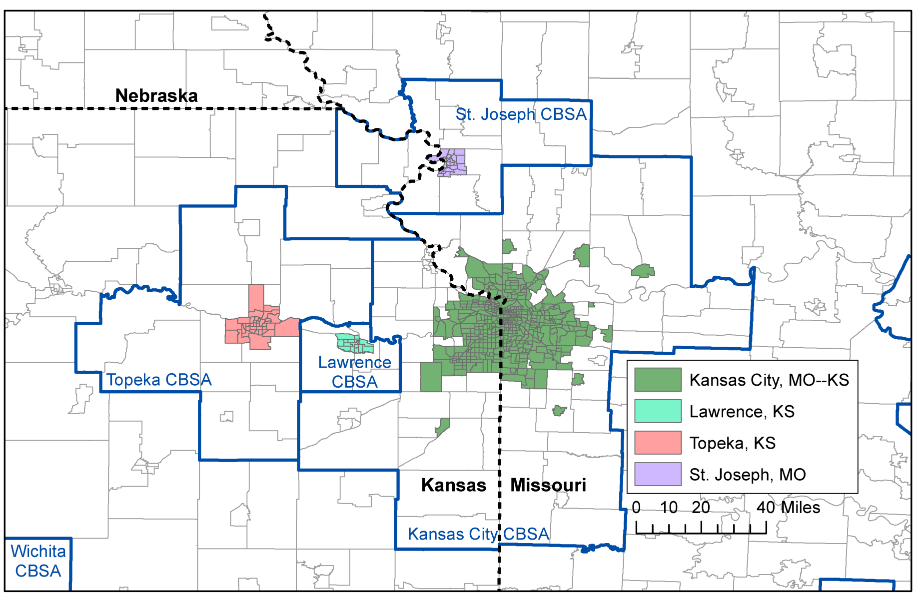
KBMAs along the Atlantic Coast of southern New England are tightly packed. For

## Figure 4: Composition of the Kansas City KBMA




Notes: Census tracts, demarcated in gray, have population density that is inversely correlated with their land area. The displayed area lies entirely within the Kansas City MO-KS metropolitan CBSA.

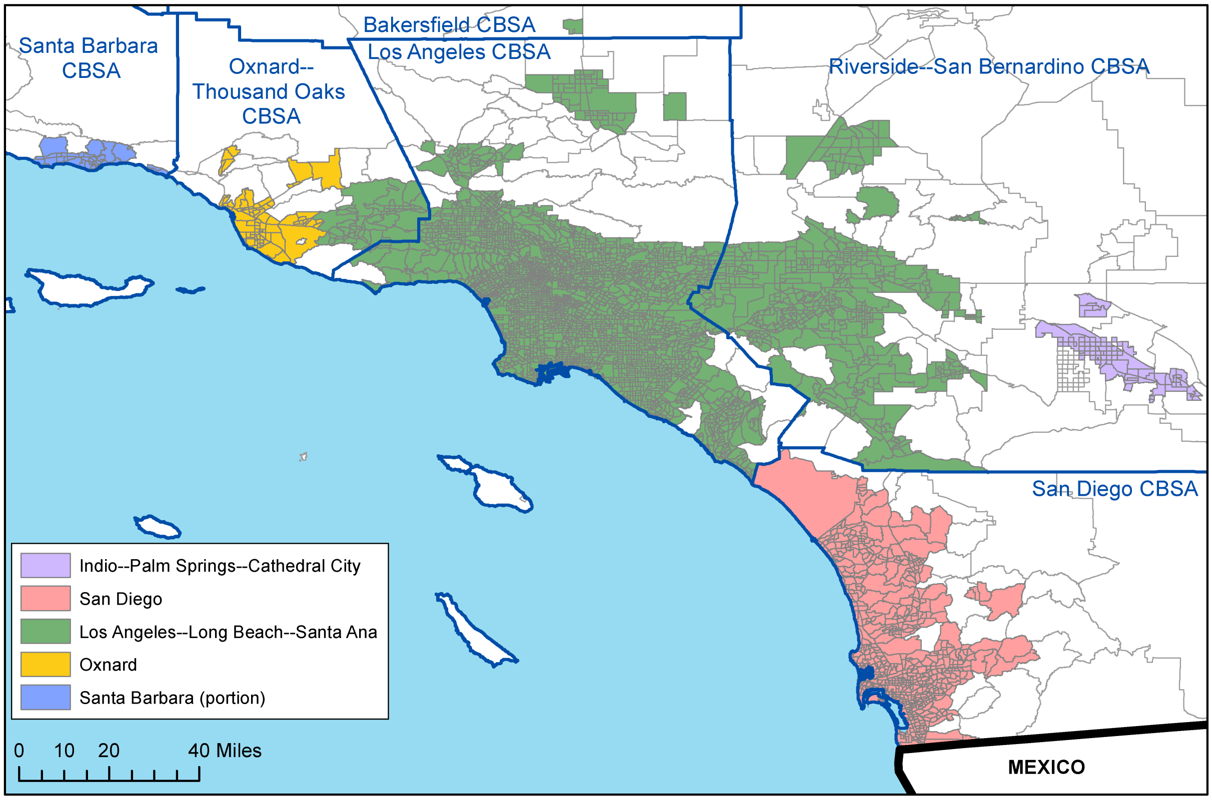
## Figure 5: Kansas City and Neighboring KBMAs.




Notes: Blue lines demarcate the borders of metropolitan CBSAs. Census tracts, demarcated in gray, have population density that is inversely correlated with their land area.

example, as illustrated in Figure 7, the Boston KBMA adjoins four other KBMAs and lies within 30 miles of four more. It also encompasses the most densely settled portions

**Figure 6: Los Angeles and Neighboring KBMAs**



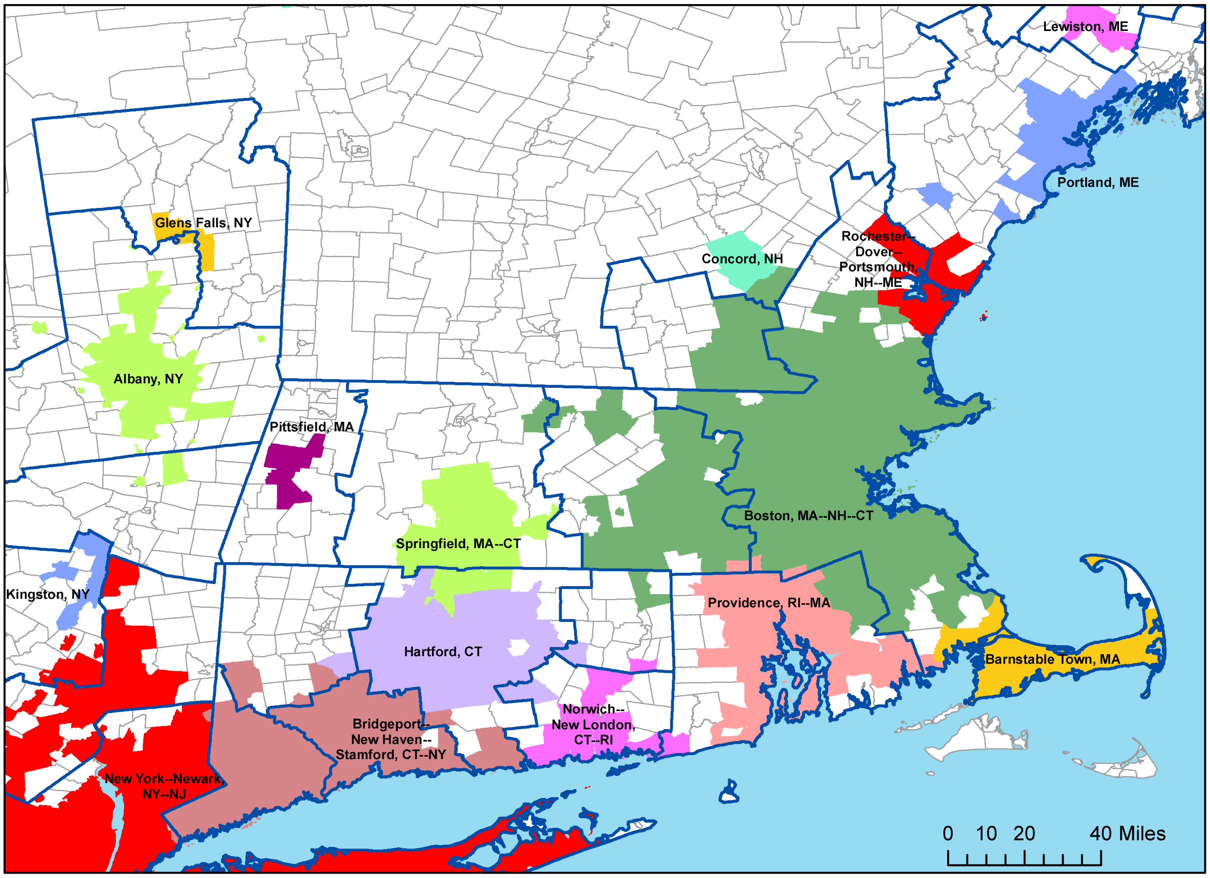

Notes: Blue lines demarcate the borders of metropolitan CBSAs. Census tracts, demarcated in gray, have population density that is inversely correlated with their land area.

of three metropolitan CBSAs: Boston, Worcester, and Manchester-Nashua.

Figure 8 shows the kernel-based urban areas in the same footprint, which fragment many KBMAs. For example, the Boston KBMA fragments into four separate kernel-based urban areas: Boston, Worcester, Leominster–Fitchburg, and Manchester–Nashua. Other KBMAs "disappear": the built-out kernels corresponding to the Kingston, Glens Falls, Pittsfield, Concord, and Lewiston KBMAs no longer have population that exceeds 50,000 and so fail to qualify as kernel-based urban areas. The median land area of the surviving 346 kernel-based urban areas is 34 percent smaller than the median land area of KBMAs.

Conversely, the kernel-based metropolitan regions in the same footprint, shown in Figure 9, consolidate many KBMAs and vastly expand them outward to encompass rural areas. For example, the Boston kernel-based metropolitan region spans five KBMAs: Boston, Providence, Barnstable Town, Rochester-Dover-Portsmouth, and Concord; its land area is more than twice the aggregate of the five KBMAs. The median land area of

**Figure 7: KBMAs in Southern New England**




Notes: Blue lines demarcate the borders of metropolitan CBSAs. Gray lines demarcate tract borders.

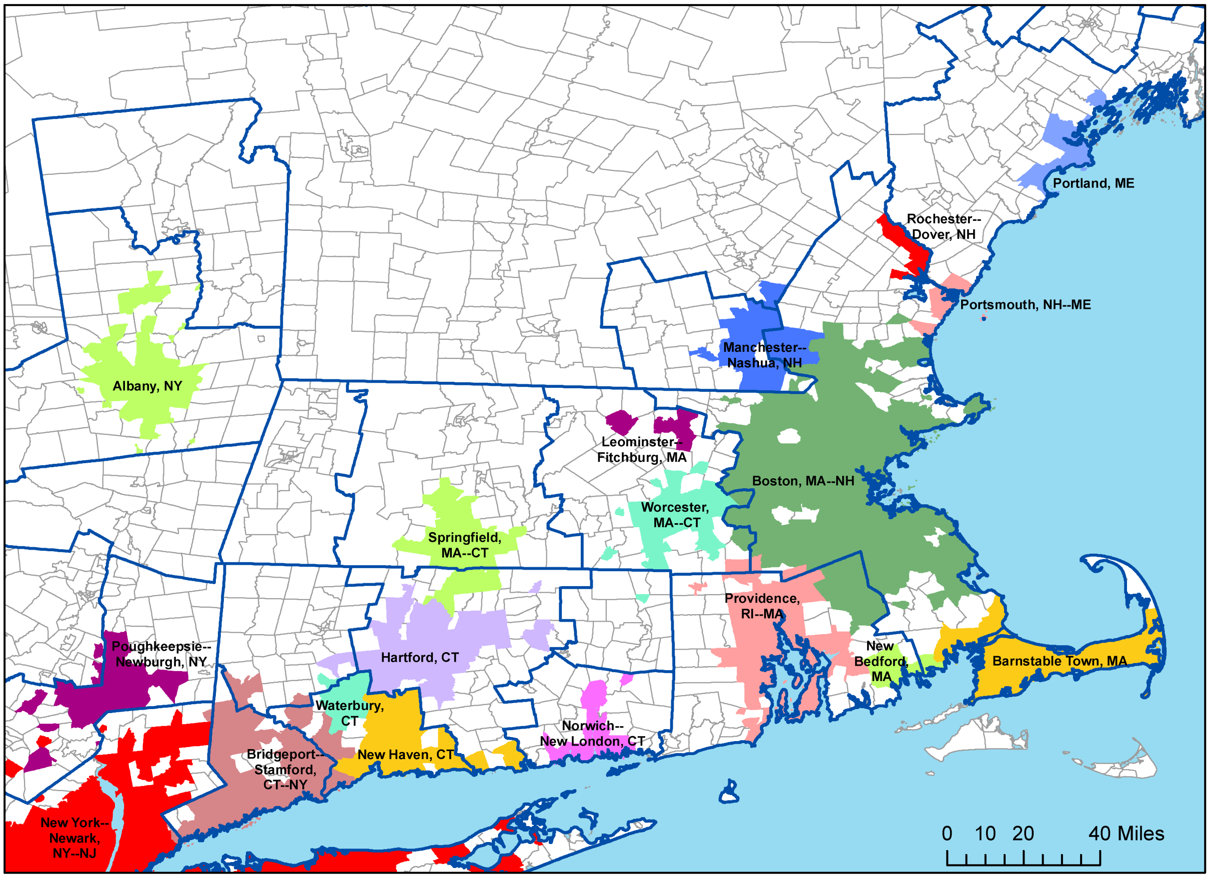
kernel-based metropolitan regions is more than 16 times that of KBMAs.

## 4.2 Comparison to Metropolitan CBSAs

Table 4 compares KBMAs with the metropolitan CBSA in which they have their most populous core. This matching is straightforward for the majority of KBMAs but less obvious for others. For example, the criterion implies comparing both the Boston and Rochester–Dover–Portsmouth KBMAs to the Boston–Cambridge–Quincy CBSA. In consequence, some metropolitan CBSAs, such as Worcester and Manchester–Nashua, have no KBMA compared against them. A complementary appendix table reciprocally compares metropolitan CBSAs with the KBMA that has the most populous core in them.

Most KBMAs have population moderately below that of their comparison CBSA,

**Figure 8: Kernel-Based Urban Areas in Southern New England**

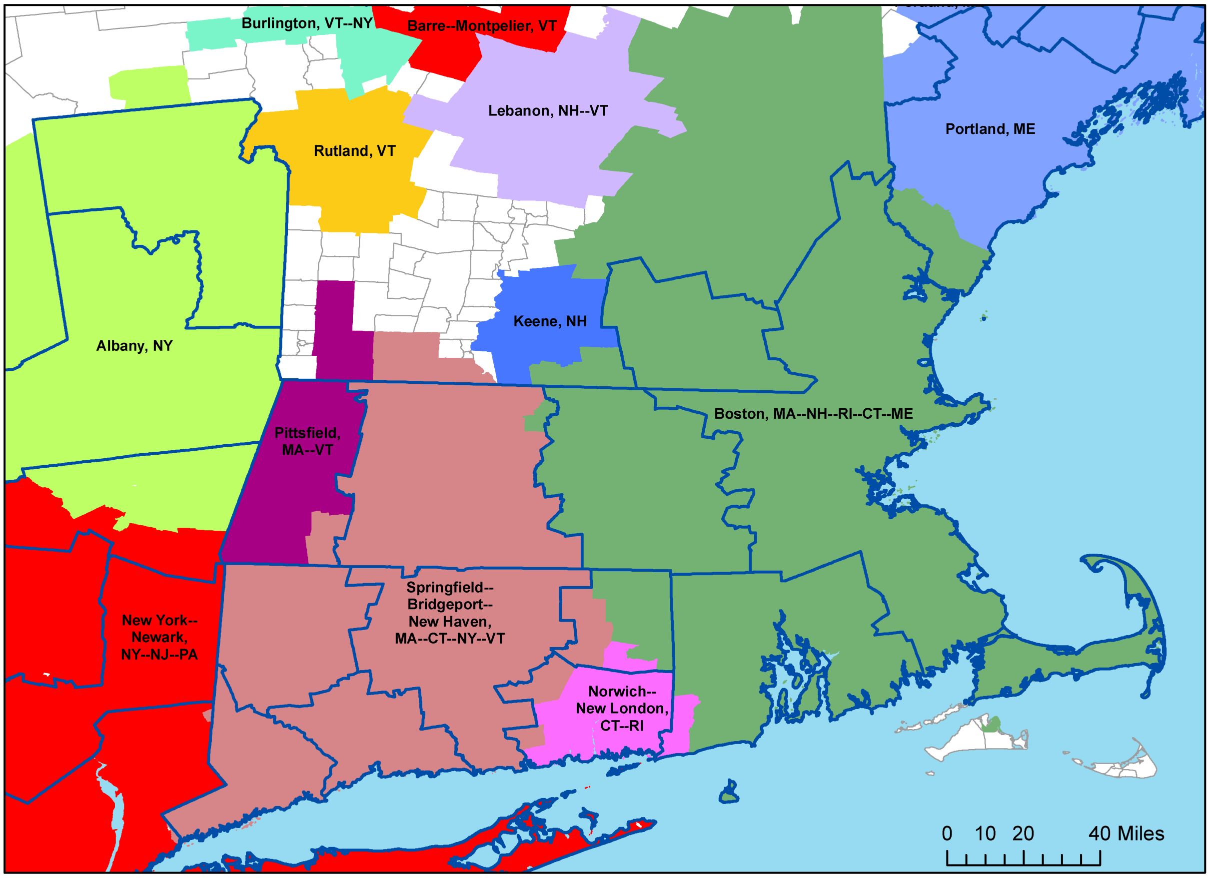



Notes: Blue lines demarcate the borders of metropolitan CBSAs. Gray lines demarcate tract borders.

employment modestly below it, and land area far below it. The median ratio of KBMA size to metropolitan CBSA size is 0.72 for population, 0.87 for employment, and 0.12 for land area. But some KBMAs that have cores in multiple metropolitan CBSAs, such as the Boston KBMA, considerably exceed their comparison CBSA in population.

Figure 10 correspondingly plots the population of KBMAs against their comparison metropolitan CBSAs; the dashed line is drawn where pairs would have equal population. The labeled KBMAs with solid markers above the dashed line have at least one core that anchors a different metropolitan CBSA. The labeled KBMAs with solid markers below the dashed line are compared to a metropolitan CBSA in which another KBMA also has a core.

**Figure 9: Kernel-Based Metropolitan Regions in Southern New England**




Notes: Blue lines demarcate the borders of metropolitan CBSAs. Gray lines demarcate tract borders.

## 4.3 Commuting

Figure 11 shows commuting inflow and outflow rates for KBMAs and metropolitan CBSAs. Measured by the enumerated median and 90th percentile rates, KBMAs underperform metropolitan CBSAs in encompassing commuting, unsurprising in the context of the KBMA parameterization's tradeoff to exclude sparsely settled land.

KBMAs are more competitive by other measures of commuting. Their flow rates have upper envelopes that decline more steeply with population, reflecting that KBMAs more successfully hold down flows for large metropolitan areas. For example, the 90th-percentile inflow rate for KBMAs with population above 500,000 is 0.20, only modestly above the 90th-percentile inflow rate of 0.17 for metropolitan CBSAs with population above 500,000. Even more competitive, the 90th percentile outflow rate for these large

Table 4: KBMAs Compared to Matched Metropolitan CBSAs

| KBMA | cores | metro CBSAs in which a core | comparison metropolitan CBSA | relative size pop | relative size emp | relative size land | share in comparison pop | share in comparison emp | share in comparison land |
|---|---|---|---|---|---|---|---|---|---|
| Kansas City, MO–KS | 3 | 1 | Kansas City, MO-KS | 0.82 | 0.90 | 0.14 | 1.00 | 1.00 | 1.00 |
| Los Angeles–Long Beach–Santa Ana | 14 | 4 | Los Angeles–Long Beach–Santa Ana | 1.23 | 1.17 | 0.83 | 0.81 | 0.85 | 0.56 |
| Indio–Palm Springs–Cathedral City | 2 | 1 | Riverside–San Bernardino–Ontario | 0.08 | 0.10 | 0.01 | 1.00 | 1.00 | 1.00 |
| Oxnard, CA | 3 | 1 | Oxnard–Thousand Oaks–Ventura | 0.55 | 0.58 | 0.12 | 1.00 | 1.00 | 1.00 |
| Boston, MA–NH–CT | 6 | 4 | Boston–Cambridge–Quincy, MA-NH | 1.20 | 1.18 | 1.14 | 0.78 | 0.80 | 0.62 |
| Rochester–Dover–Portsmouth,NH-ME | 2 | 2 | Boston–Cambridge–Quincy, MA–NH | 0.04 | 0.04 | 0.08 | 0.83 | 0.87 | 0.72 |
| Providence, RI–MA | 2 | 3 | Providence-New Bedford-Fall River, RI-MA | 0.88 | 0.90 | 0.60 | 0.98 | 0.99 | 0.95 |
| minimum | 1 | 0 | | 0.01 | 0.01 | 0.00 | 0.30 | 0.40 | 0.20 |
| 10th percentile | 1 | 1 | | 0.49 | 0.64 | 0.03 | 0.88 | 0.92 | 0.82 |
| median | 1 | 1 | | 0.72 | 0.87 | 0.12 | 1.00 | 1.00 | 1.00 |
| 90th percentile | 3 | 2 | | 0.97 | 0.99 | 0.39 | 1.00 | 1.00 | 1.00 |
| maximum | 14 | 6 | | 2.09 | 1.95 | 2.26 | 1.00 | 1.00 | 1.00 |
| aggregate (all KBMAs relative to all metropolitan CBSAs) | | | | 0.88 | 0.94 | 0.15 | 0.99 | 0.99 | 0.96 |

Notes: Each KBMA is compared to the metropolitan CBSA in which it has its most populous core. The summary statistics on relative size and overlap exclude the 19 KBMAs whose largest core is in a micropolitan CBSA. The aggregate statistics compare all KBMAs with all metropolitan CBSAs. A complementary appendix table compares metropolitan CBSAs with the KBMA that has the most populous core in them.

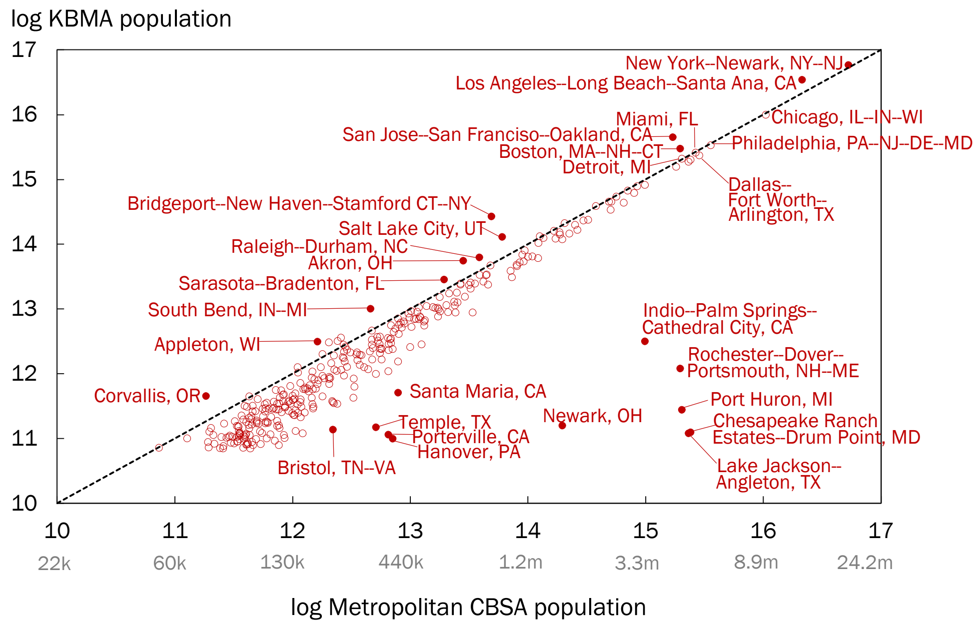
KBMAs is 0.13, below the outflow rate of 0.17 for the large metropolitan CBSAs.

As illustrated in Figure 12, the kernel-based urban areas have commuting inflow and outflow rates moderately higher than those of KBMAs. Conversely, the kernel-based metropolitan regions have flow rates that are considerably lower. As illustrated in an appendix figure, the kernel-based urban area flow rates are similar to those of Urbanized Areas, and the kernel-based metropolitan region flow rates are similar to those of Commuting Zones.

## 4.4 Size

Figure 13 shows the histogram of KBMA population: on the left compared against those of metropolitan CBSAs and UAs; on the right, against those of kernel-based metropolitan regions and kernel-based urban areas. The KBMA distribution slopes downward across all of the bins, approximately matching the UA distribution. In contrast, the metropolitan CBSA distribution has frequency that increases across the three lowest bins

**Figure 10: Population of KBMAs versus Comparison Metropolitan CBSA**

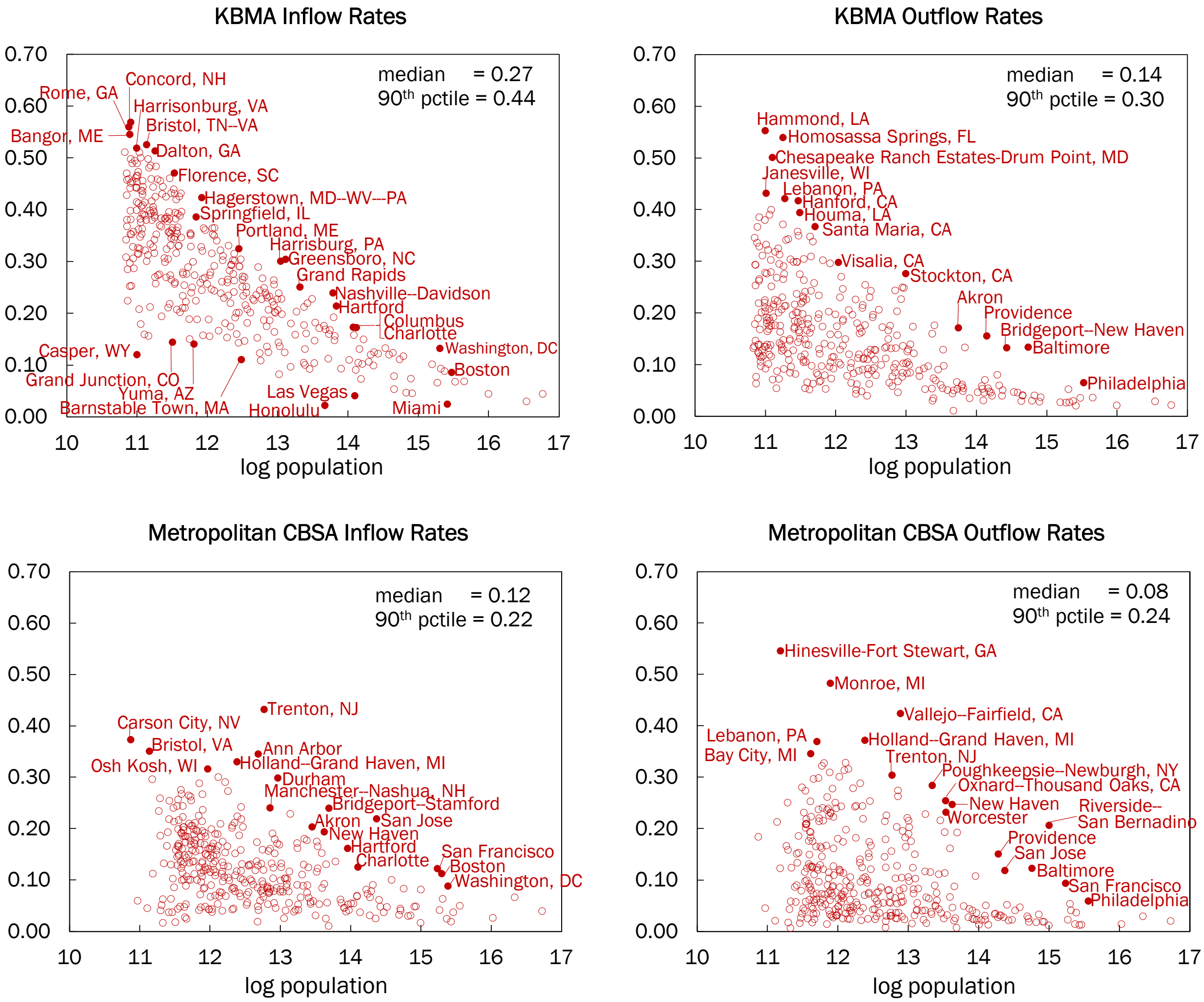



Notes: Each KBMA is compared to the metropolitan CBSA in which it has its most populous core. The dashed line is where KBMA and metropolitan CBSA population equal each other. Labeled KBMAs with solid markers above the dashed line include at least one core that anchors a different metropolitan CBSA. Labeled KBMAs with solid markers below the dashed line are compared to a metropolitan CBSA in which another KBMA also has a core.

before turning down. The upward-sloping portion reflects that the scale criterion for metropolitan CBSAs applies to their core population rather than total population. To the extent that population buildouts—the ratio of non-core to core population—have a unimodal distribution, relatively few metropolitan CBSAs will have population only moderately above 50,000. In contrast, the more cleanly truncated set of all CBSAs with population above 50,000, both metropolitan and micropolitan, has population frequency that slopes monotonically down, approximately matching the KBMA and UA distributions.

As illustrated in the right panel, the population distributions of the three kernel-based parameterizations approximately match each other. This may seem counterintuitive: During the kernel construction stage, the additional iterations associated with a lower threshold commuting strength and a higher allowed separating distance merge together what otherwise would be separate kernels, shifting the population distribution of built-out kernels rightward. But the shift also pushes more built-out kernels above the 50,000

**Figure 11: Commuting Flows: KBMAs and Metropolitan CBSAs**

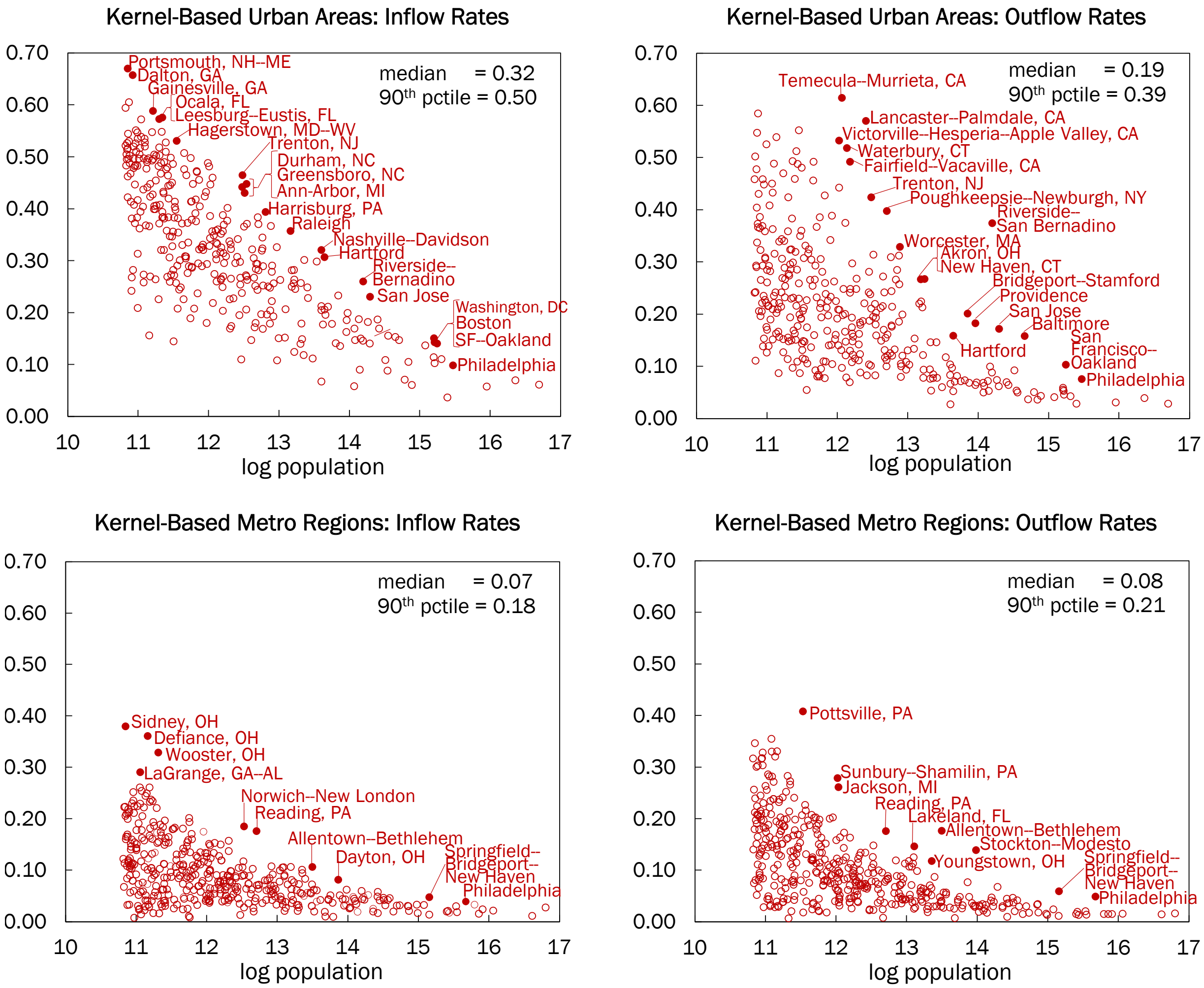



Notes: Inflow rates are measured as a share of employment; outflows rates, as a share of residents who are employed.

threshold, contributing to keeping the population distribution approximately unchanged. Equally important, as described previously, the iterative calculation of pairwise commuting strength avoids constructing long chains of large cores, limiting the rightward shift at the top of the population distribution. In aggregate, the 435 kernel-based metropolitan regions have 29 percent more population than the 361 KBMAs; the 346 kernel-based urban areas have 9 percent less population.

Figure 14 plots the logarithm of the population rank of observations for each of the kernel-based delineations against the logarithm of their population, a standard benchmark for local delineations (e.g., Rosen and Resnick, 1980; Gabaix, 1999; Gabaix and

**Figure 12: Commuting Flows: Kernel-Based Metropolitan Regions and Kernel-Based Urban Areas**



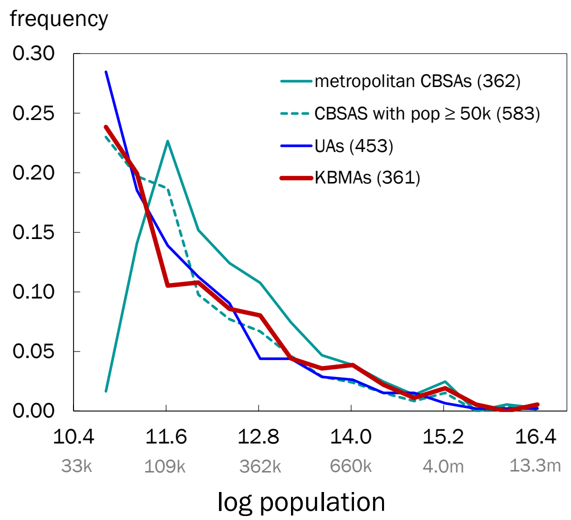

Notes: Inflow rates are measured as a share of employment; outflow rates, as a share of residents who are employed. The Tracy, CA kernel-based urban area, which has log population of 10.9, has a commuting outflow rate of 0.76, above the displayed range.

Ioannides, 2004; Soo, 2005; Rozenfeld et al., 2011). The logarithmic rank–size relationship is expected to be approximately linear for Pareto distributions, with a slope close to -1 if the distribution is Zipf, a specialization of Pareto (Gabaix, 1999).[11]

Instead, the plots are visibly concave, notwithstanding estimated linear relationships with R-squared values very close to 1. As illustrated in the appendix, the same combination of visible concavity and R-squared values close to 1 for linear estimates describe

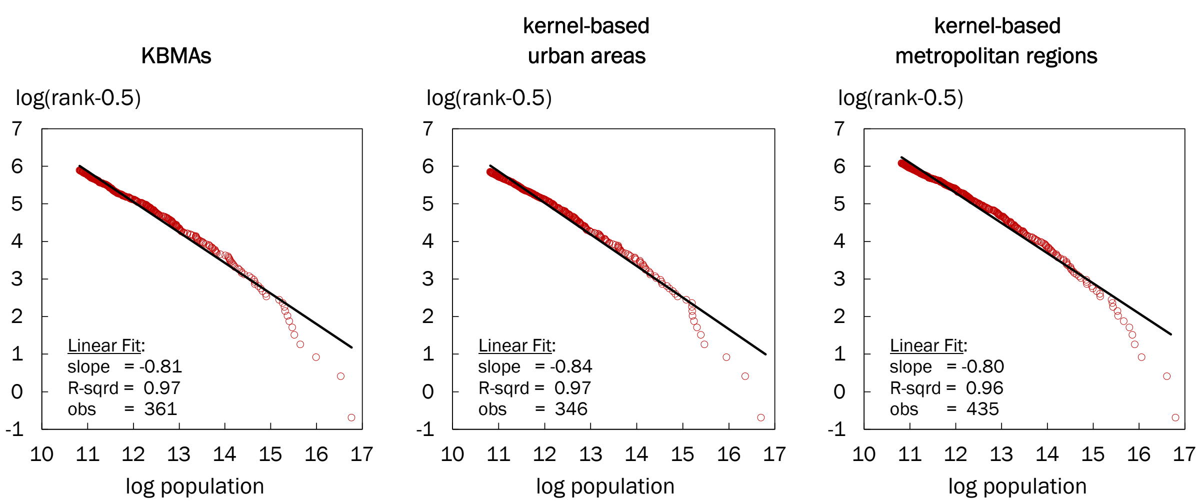
[11] A Pareto distribution has CDF, $F(S) = 1 - (\widetilde{S}/S)^{\zeta}$ for size $S \geq \widetilde{S}$, $\zeta > 0$. It specializes to a Zipf's distribution when the shape parameter, $\zeta$, equals 1.

## Figure 13: The Distribution of Metropolitan Population




Notes: The horizontal axes measure the logarithm of population at the lower bound of bins with width of 0.4 log points. The lower bound of the leftmost bin corresponds to a population of 50,000 for all histograms. The green-dashed histogram in the left panel describes the set of all CBSAs with population of at least 50,000, both metropolitan and micropolitan.

## Figure 14: Population Rank versus Size




Note: Subtracting 0.5 from the rank improves the fit in the presence of small-sample bias (Gabaix and Ibragimov, 2011).

analogous plots for UAs and metropolitan CBSAs. The apparent contradiction reflects peculiarities of the logarithmic rank-size relationship (Eeckhout, 2009). In particular, the visual relationship is distorted across the most populous locations, reflecting the large differences in logarithmic values across the smallest integer ranks. The high R-squared values are also misleading, among other reasons because the relationship is monotonic by construction. Consistent with this, Monte Carlo simulations establish that estimated standard errors from OLS regressions grossly understate stochastic uncertainty (Gabaix and Ioannides, 2004). As reported in an appendix table, we thus *cannot* reject that the largest 50 observations of each of the kernel-based delineations are drawn from a Pareto distribution.

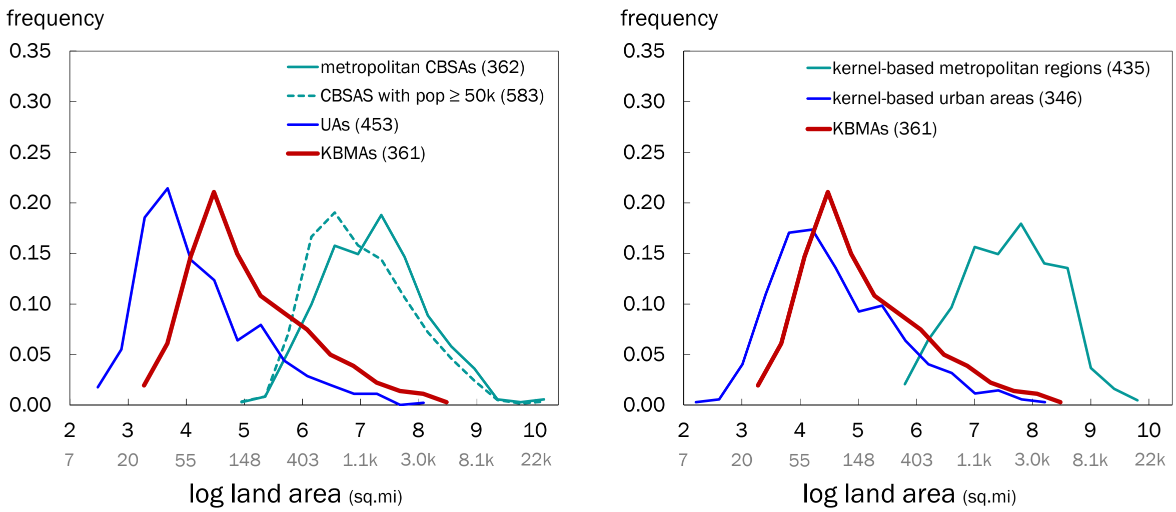
In contrast, the Monte Carlo simulations do reject Pareto at the 0.05 level for the largest 75 observations of each of the kernel-based delineations as well as for the largest 75 UAs and metropolitan CBSAs. They also reject Pareto across a much wider range of kernel-based parameterizations, including permutations that separately parameterize the threshold commuting strength and maximum separating distance at the kernel and buildout stages of construction (Rappaport and Humann, 2023).

A benchmark alternative hypothesis is that population distributions are lognormal over their entire range of observations, spanning from very low levels of population to the largest. For example, Eeckhout (2004) shows that the union of municipalities and census designated places in 2000, which includes locations with population as low as 1, is well approximated by a lognormal population distribution with peak frequency at a population close to 1,400.[12] One reason lognormal is important is that it is asymptotically implied by stochastic processes that satisfy Gibrat's law, having growth rates that are uncorrelated with initial levels (Eeckhout, 2004; Lee and Li, 2013).

We are skeptical. As illustrated in the appendix, the frequency distribution of the union of UAs and UCs is convex downward sloping from its minimum population of 2,500. The 3,609 UAs and UCs in 2000 together accounted for only 79 percent of the U.S. population, at minimum requiring tens of thousands of smaller clusters to encompass the remaining 59 million residents. Similarly, the union of municipalities and census designated places in Eeckhout (2004) accounted for only 71 percent of the U.S. population (Levy, 2009). Bolstering our skepticism, more recent research finds that population growth

[12]Census designated places are statistical entities consisting of closely settled, locally recognized concentrations of population that are identified by name. Designation relies heavily on the opinions of local residents, organizations, and officials (U.S. Census Bureau, 1997*a*). This reliance induces a selection bias: local support for designation is likely to be stronger for locations with larger population.

**Figure 15: Distribution of Metropolitan Land Area**



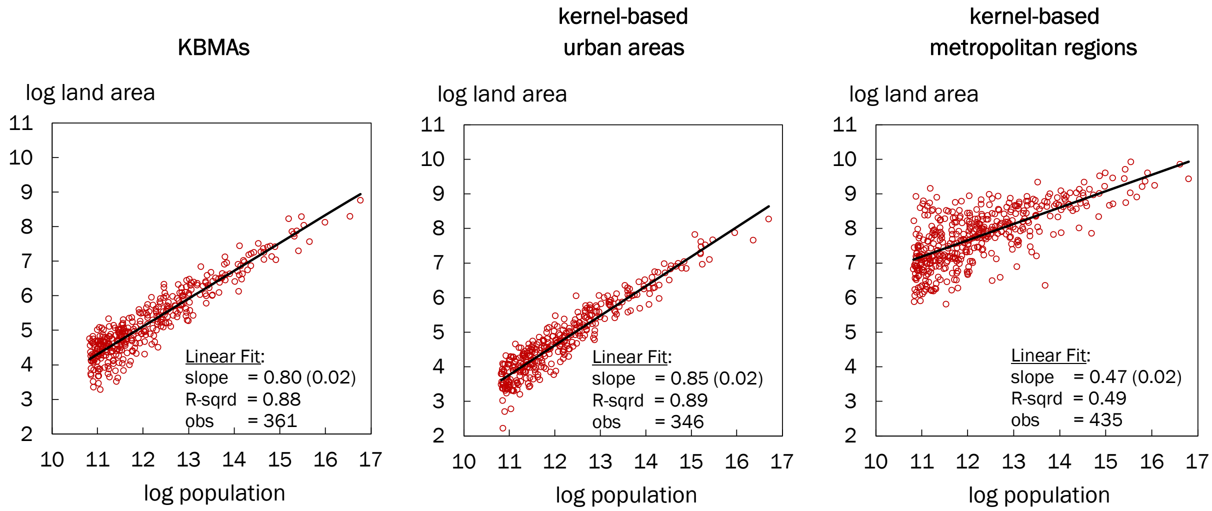

Notes: The horizontal axes measure the logarithm of land area of the lower bound of bins with width of 0.4 log points. For each histogram, the lower bound of the leftmost bin corresponds to the observation with minimum land area.

rates of various geographical delineations were historically correlated with population levels (Holmes and Lee, 2010; Dittmar, 2021; Michaels, Rauch and Redding, 2012; Desmet and Rappaport, 2017; Rappaport, 2018). For example, Rappaport (2018) documents that locations with intermediate levels of population grew fastest during recent decades.

Figure 15 compares histograms of land area. Unsurprisingly, the KBMA land distribution is located to the right of the UA distribution and considerably to the left of the metropolitan CBSA distribution. Less obviously, the KBMA distribution represents a near parallel rightward shift from the UA distribution, including preserving the latter's pronounced rightward skew. Similarly, the land distribution of kernel-based urban areas, shown in the right panel, is a parallel shift rightward from UAs, but by less far than that of KBMAs. In aggregate, the 346 kernel-based urban areas have 39 percent less land area than the 361 KBMAs.

The land distribution of kernel-based metropolitan regions compresses that of metropolitan CBSAs, including truncating the latter's right tail. For example, the Dallas–Fort Worth–Arlington kernel-based metropolitan region has the largest land area, 20,500 square miles, more than 20 percent below the land areas of the Riverside-San Bernadino and Anchorage metropolitan CBSAs. The least expansive metropolitan region occupies

**Figure 16: Land Area versus Population**




Note: Standard errors for the slope coefficient are reported in parentheses.

334 square miles, more than twice that of the least expansive metropolitan CBSA. In aggregate, the 435 kernel-based metropolitan regions have land area more than 10 times that of the 361 KBMAs.

Figure 16 plots the logarithm of land area against the logarithm of population for the three kernel-based delineations. The relationship is tight for both KBMAs and the kernel-based urban areas, with respective estimated coefficients of 0.80 and 0.85 and R-squared values close to 0.90. The relationship for kernel-based metropolitan regions is shallower and less tight, with an estimated coefficient of 0.47 and an R-squared value of 0.49. As illustrated in the appendix, the estimated relationship for the kernel-based urban areas almost exactly matches that for UAs, and the estimated coefficient for the kernel-based metropolitan regions is similar to that for metropolitan CBSAs but with a tighter fit. Our estimates are also squarely within the range reported in previous research: Ahlfeldt and Pietrostefani (2019) report a coefficient of 0.57 estimated across 70 U.S. Functional Urban Areas, and analogous coefficients ranging from 0.29 to 0.92 across Functional Urban Areas in seven additional countries for which they have at least 10 observations. Similarly, Combes, Duranton and Gobillon (2019) report an estimate of about 0.7 across urban areas in France.

Estimated coefficients for such logarithmic plots correspond to the implicit elasticity of land area with respect to population. All are well below 1, implying that relative differences in land area are less than proportional to relative differences in population. This less-than-proportional relationship suggests centripetal forces pull residents and establishments to locate near central locations, for example to hold down the commuting cost of accessing clustered employment (Alonso, 1964; Mills, 1967; Muth, 1969).[13]

Consistent with centripetal forces, the relationship of land to population for the union of UAs and UCs is concave, statistically significant at the 0.01 level. As illustrated in an appendix figure, the fitted elasticity declines from 1.01 at the smallest observation (population = 2,500) to 0.71 at the largest (population = 18.3 million). This concavity probably also reflects topography, which constrains geographic expansion in many large metropolitan areas (Saiz, 2010).

Importantly, the differences in the estimated elasticities for the three kernel-based delineations more likely reflect the capacity to accommodate increases in population within a fixed footprint rather than structural barriers to expansion. The estimated elasticity is lowest for the kernel-based metropolitan regions: by construction, they encompass vast amounts of rural land and so can accommodate large increases in population without expanding their footprint. The estimated elasticity is highest for the kernel-based urban areas: by construction, they are the most built-up and so probably have the lowest capacity to accommodate increases in population without expanding.

# 5 Conclusion

Parameterizations of our kernel-based algorithm better match a broad definition of metropolitan areas compared to metropolitan CBSAs and other existing delineations. The KBMA parameterization, which balances encompassing commuting flows and excluding sparsely settled land, is likely to be appropriate for most purposes, including for planning, regulation, and research. The more expansive parameterization for kernel-based

[13]For example, Rappaport (2016) models metropolitan population and land area as endogenously depending on exogenous differences in business productivity. Land area proportionally increases almost one-to-one with (productivity-driven) increases in population from low levels but by considerably less with increases in population from high levels.

metropolitan regions or the more compact parameterization for kernel-based urban areas may be appropriate for other purposes. Datasets for each are included in the paper's supplemental materials, and a replication package allows users to customize delineations as they judge appropriate (Rappaport and Humann, 2025*b*,*a*).

A priority for future research is constructing kernel-based delineations for more recent years. Doing so will realize another advantage of KBMAs compared to CBSAs, the ability to decompose metropolitan growth into intensive and extensive margins (i.e., within and outside starting footprints). CBSAs poorly do so because settlement varies hugely within its county building blocks. Similarly, kernel-based delineations can closely track the geographic expansion of nearby metropolitan areas up against each other, in some cases merging into a single metropolitan area but not in others.

On the other hand, using census tracts as building blocks limits data availability. To be sure, the Census Bureau publishes tabulated tract data for some of its products, including decennial censuses and the American Community Survey. Conversely, anonymity requirements limit the availability of microdata for most counties and metropolitan CBSAs, lessening the disadvantage of using census tracts as building blocks.

Duranton (2021) argues that appropriate metropolitan delineations are required to address fundamental urban questions. We hope that KBMAs and other parameterizations of our kernel-based algorithm will contribute to doing so.

# Additional Materials

The supplemental materials package, Rappaport and Humann (2025*b*), includes enumerations of KBMAs, kernel-based metropolitan regions, and kernel-based urban areas; several workbooks for each with detailed variables of their components (census tracts, cores, kernels, and built-out kernels); maps for each; shape files for each; the sequence of iterative joins down to a strength of 0.01 for each of four distance maximums; and pairwise commuting flows between metropolitan CBSAs. The package's individual files are separately available from the working paper's landing page on the website of the Federal Reserve Bank of Kansas City.

The replication package, Rappaport and Humann (2025*a*), includes Matlab files for constructing kernel-based areas, including "start" files with the parameterizations to construct KBMAs, kernel-based metropolitan regions, and kernel-based urban areas; python files to generate titles for each of the three parameterizations; all required data files; CSV files of all constructed components for each of the three parameterizations; and a readme file with a description of how to modify the start files to construct custom parameterizations.

# References


**Ahlfeldt, Gabriel M., and Elisabetta Pietrostefani.** 2019. "Appendix to The Economic Effects of Density: a Synthesis." *Journal of Urban Economics.*

**Alonso, William.** 1964. *Location and Land Use.* Cambridge:Harvard University Press.

**Arribas-Bel, Daniel, Miquel-Angel Garcia-Lopez, and Elisabet Viladecans-Marsal.** 2021. "Building(s and) Cities: Delineating Urban Areas with a Machine Learning Algorithm." *Journal of Urban Economics*, 125.

**Autor, David, David Dorn, and Gordon Hanson.** 2013. "The China Syndrome: Local Labor Market Effects of Import Competition in the United States." *American Economic Review*, 103: 2121–2168.

**Berry, Brian J.L., Peter G. Goheen, and Harry Goldstein.** 1969. "Metropolitan Area Definition: A Re-Evaluation of Concept and Statistical Practice." Bureau of Census.

**Brezzi, Monica, Mario Piacenti, Konstantin Rosina, and Daniel Sanchez-Serra.** 2012. "Redefining Urban Areas in OECD Countries." In *Redefining "Urban": A New Way to Measure Metropolitan Areas.* Chapter 1. Paris:OECD Publishing.

**Bureau of Transportation Statistics.** 2005. "Census Transportation Planning Package (CTPP) 2000."

**Chetty, Raj, Nathaniel Hendren, Patrick Kline, and Emmanuel Saez.** 2014. "Where it the land of Opportunity? The Geography of Intergenerational Mobility in the United States." *Quarterly Journal of Economics*, 129: 1553–1623.

**Ch, Rafael, Diego A. Martin, and Juan F. Vargas.** 2021. "Measuring the Size and Growth of Cities Using Nighttime Light." *Journal of Urban Economics*, 125.

**Combes, Pierre-Philippe, Gilles Duranton, and Laurent Gobillon.** 2008. "Spatial Wage Disparities: Sorting Matters!" *Journal of Urban Economics*, 62: 723–742.

**Combes, Pierre-Philippe, Gilles Duranton, and Laurent Gobillon.** 2019. "The Costs of Agglomeration: House and Land Prices in French Cities." *Review of Economic Studies*, 86: 1556–1589.

**Congressional Research Service.** 2014. "Metropolitan Area Designations by OMB: History, 2010 Standards, and Uses."

**de Bellefon, Marie-Pierre, Pierre-Philippe Combes, Gilles Duranton, Laurent Gobillon, and Clément Gorin.** 2021. "Delineating Urban Areas Using Building Density." *Journal of Urban Economics*, 125.

**Desmet, Klaus, and Jordan Rappaport.** 2017. "The Settlement of the United States, 1800-2000: The Long Transition towards Gibrat's Law." *Journal of Urban Economics*, 98: 50–68.

**Dijkstra, Lewis, and Hugo Poelman.** 2012. "Cities in Europe: The New OECD-EC Definition." European Commision, Directorate-General for Regional and Urban Policy.

**Dijkstra, Lewis, Aneta J. Florczyk, Sergio Freire, Thomas Kemper, Michele Melchiorri, Martino Pesaresi, and Marcello Schiavina.** 2021. "Applying the Degree of Urbanisation to the Globe: a New Harmonised Definition Reveals a Different Picture of Global Urbanisation." *Journal of Urban Economics*, 125.

**Dijkstra, Lewis, Hugo Poelman, and Paolo Veneri.** 2019. "The EU-OECD Definition of a Functional Urban Area." OECD.

**Dingel, Jonathan I., Antonio Miscio, and Donald R. Davis.** 2021. "Cities, Lights, and Skills in Developing Economies." *Journal of Urban Economics*, 125.

**Dittmar, Jeremiah.** 2021. "Cities, Markets, and Growth: The Emergence of Zipf's Law." Institute for Advanced Study.

**Duranton, Gilles.** 2015. "Delineating Metropolitan Areas: Measuring Spatial Labour Market Networks Through Commuting Patterns." In *The Economics of Interfirm Networks.* , ed. Iichiro Uesugi Tsutomu Watanabe and Arito Ono, 107–133. Tokyo:Springer.

**Duranton, Gilles.** 2021. "Classifying Locations and Delineating Space: An Introduction." *Journal of Urban Economics*, 125.

**Economic Research Service.** 2012. "2000 Commuting Zones." data retrieved on 3/24/2020 from https://www.ers.usda.gov/webdocs/DataFiles/48457/cz00_eqv_v1.xls.

**Eeckhout, Jan.** 2004. "Gibrat's Law for (All) Cities." *American Economic Review*, 94: 1429–1451.

**Eeckhout, Jan.** 2009. "Gibrat's Law for (All) Cities: Reply." *American Economic Review*, 99: 1676–1683.

**Eurostat, OECD, United Nations, and The World Bank.** 2021. *Applying the Degree of Urbanisation: A Methodological Manual to Define Cities, Towns and Rural Areas for International Comparisons.* Paris and Brussels:OECD Publishing.

**Federal Financial Institutions Examination Council.** 2024. "Median Family Income Report." data retrieved on 1/31/2025 from https://www.ffiec.gov/sites/default/files/data/media-income/msa24inc.xls.

**Foote, Andrew, Mark J. Kutzbach, and Lars Vilhuber.** 2017. "Recalculating...: How Uncertainty in Local Labor Market Definitions Affects Empirical Findings." U.S. Census Bureau, Center for Economic Studies.

**Gabaix, Xavier.** 1999. "Zipf's Law for Cities: an Explanation." *Quarterly Journal of Economics*, 114.

**Gabaix, Xavier, and Rostam Ibragimov.** 2011. "Rank - 1/2: A Simple Way to Improve the OLS Estimation of Tail Exponents." *Journal of Business and Economic Statistics.*

**Gabaix, Xavier, and Yannis M. Ioannides.** 2004. "The Evolution of City Size Distributions." In *Handbook of Regional and Urban Economics.* , ed. J.V Henderson and J.F. Thisse, 173–204. Amsterdam:North Holland.

**Galdo, Virgilio, Yue Li, and Martin Rama.** 2021. "Identifying Urban Areas by Combining Human Judgement and Machine Learning: An Application to India." *Journal of Urban Economics*, 125.

**Gardner, Todd.** 1999. "Metropolitan Classification for Census Years before World War II." *Historial Methods*, 32.

**Gardner, Todd.** 2021. "Changes in Metropolitan Area Definition, 1910-2010." U.S. Census Bureau.

**Glaeser, Edward L., and David C. Mare.** 2001. "Cities and Skills." *Journal of Labor Economics*, 19: 316–342.

**Green, Richard K., Stephen Malpezzi, and Stephen K. Mayo.** 2005. "Metropolitan-Specific Estimates of the Price Elasticity of the Supply of Housing and Their Sources." *American Economic Review*, 95: 334–339.

**Greenstone, Michael, Richard Hornbeck, and Enrico Moretti.** 2010. "Identifying Agglomeration Spillovers: Evidence from Winners and Losers of Large Plant Openings." *Journal of Political Economy*, 118: 536–598.

**Guren, Adam M., Alisdair McKay, Emi Nakamura, and Jón Steinson.** 2021. "Housing Wealth Effects." *The Review of Economic Studies*, 88: 669–707.

**Holmes, Thomas J., and Sanghoon Lee.** 2010. "Cities as Six-by-Six-Mile Squares: Zipf's Law?" In *Agglomeration Economics.* , ed. Edward L. Glaeser, 105–131. Chicago and London:University of Chicago Press.

**Kennan, John, and James R. Walker.** 2011. "The Effect of Expected Income on Individual Migration Decisions." *Econometrica*, 79: 211–251.

**Landvoigt, Tim, Monika Piazzesi, and Martin Schneider.** 2015. "The Housing Market(s) of San Diego." *American Economic Review*, 105: 1371–1407.

**Lee, Sanghoon, and Qiangi Li.** 2013. "Uneven Landscapes and City Size Distributions." *Journal of Urban Economics*, 78: 19–29.

**Levy, Moshe.** 2009. "Gibrat's Law for (All) Cities: Comment." *American Economic Review*, 99: 1672–1675.

**Mian, Atif, and Amir Sufi.** 2012. "The Effects of Fiscal Stimulus: Evidence from the Cash for Clunkers Program." *Quarterly Journal of Economics*, 127: 1107–1142.

**Mian, Atif, and Amir Sufi.** 2014. "What Explains the 2007-2009 Drop in Employment?" *Econometrica*, 82: 2197–2223.

**Michaels, Guy, Ferdinand Rauch, and Stephen J. Redding.** 2012. "Urbanization and Structural Transformation." *Quarterly Journal of Economics*, 127: 535–586.

**Mills, Edwin S.** 1967. "An Aggregative Model of Resource Allocation in a Metropolitan Area." *American Economic Review*, 57: 197–210.

**Monte, Ferdinando, Stephen J. Redding, and Esteban Rossi-Hansberg.** 2018. "Commuting, Migration, and Local Employment Densities." *Amerian Economic Review*, 108: 3855–3890.

**Moretti, Enrico.** 2004. "Workers' Education, Spillovers, and Productivity: Evidence from Plant-Level Production Functions." *American Economic Review*, 94: 656–690.

**Muth, Richard F.** 1969. *Cities and Housing.* Chicago:University of Chicago Press.

**Office of Management and Budget.** 1990. "Revised Standards for Defining Metropolitan Areas in the 1990's; Notice." *Federal Register: The Daily Journal of the United States*, 55: 12154–12160.

**Office of Management and Budget.** 1998. "Alternative Approaches to Defining Metropolitan and Nonmetropolitan Areas; Notice." *Federal Register: The Daily Journal of the United States*, 63.

**Office of Management and Budget.** 2000. "Standards for Defining Metropolitan and Micropolitan Statistical Areas; Notice." *Federal Register: The Daily Journal of the United States*, 65: 82228–82238.

**Office of Management and Budget.** 2003. "Metroplitan Statistical Areas, Micropolitan Statistical Areas, Combined Statistical Areas, New England City and Town Areas, Combined New England City and Town Areas, 2003." *OMB Bulletin No. 03-04*, 82228–82238.

**Office of Management and Budget.** 2010. "2010 Standards for Delineating Metropolitan and Micropolitan Statistical Areas; Notice." *Federal Register: The Daily Journal of the United States*, 75.

**Office of Management and Budget.** 2021. "2020 Standards for Delineating Metropolitan and Micropolitan Statistical Areas." *Federal Register: The Daily Journal of the United States*, 86.

**Rappaport, Jordan.** 2016. "Productivity, Congested Commuting, and Metro Size." Federal Reserve Bank of Kansas City Working Paper.

**Rappaport, Jordan.** 2018. "The Faster Growth of Larger, Less Crowded Locations." *Economic Review*, 105. Federal Reserve Bank of Kansas City.

**Rappaport, Jordan, and McKenzie Humann.** 2023. "The Size of U.S. Metropolitan Areas." Federal Reserve Bank of Kansas City Working Paper.

**Rappaport, Jordan, and McKenzie Humann.** 2025*a*. "A Better Delineation of U.S. Metropolitan Areas: Replication Package." Mendeley Data, V1, doi: 10.17632/jht7pgb563.1 (https://data.mendeley.com/datasets/jht7pgb563/1).

**Rappaport, Jordan, and McKenzie Humann.** 2025*b*. "A Better Delineation of U.S. Metropolitan Areas: Supplemental Materials." Mendeley Data, V1, doi: 10.17632/pg64w4hkjh.1 (https://data.mendeley.com/datasets/pg64w4hkjh/1).

**Ratcliffe, Michael.** 2015. "A Century of Delineating a Changing Landscape: The Census Bureau's Urban and Rural Classification, 1910 To 2010." accessed 2/16/2022 from http://www2.census.gov/geo/pdfs/reference/ua/Century_of_Defining_Urban.pdf.

**Rosen, Kenneth T., and Mitchel Resnick.** 1980. "The Size Distribution of Cities: An Examination of the Pareto Law and Primacy." *Journal of Urban Economics*.

**Rozenfeld, Hernan D., Diego Rybski, Xavier Gabaix, and Hernan A. Makse.** 2011. "The Area and Population of Cities: New Insights from a Different Perspective on Cities." *American Economic Review*, 101: 2205–2225.

**Saiz, Albert.** 2010. "The Geographic Determinants of Housing Supply." *Quarterly Journal of Economics*, 125: 1253–1296.

**Soo, Kwok Tong.** 2005. "Zipf's Law for Cities: A Cross Country Investigation." *Regional Science and Urban Economics*, 35: 239–263.

**Tolbert, Charles M., and Molly Sizer.** 1987. "US Commuting Zones and Labor Market Areas." United States Department of Agriculture.

**Tolbert, Charles M., and Molly Sizer.** 1996. "US Commuting Zones and Labor Market Areas: a 1990 update." United States Department of Agriculture.

**U.S. Census Bureau.** 1997*a*. "Census Designated Place (CDP), Program for Census 2000–Final Criteria; Notice." *Federal Register: The Daily Journal of the United States*, 62: 44638–44641.

**U.S. Census Bureau.** 1997*b*. "Census Tract Program for Census 2000–Final Criteria; Notice." *Federal Register: The Daily Journal of the United States*, 62: 55778–55781.

**U.S. Census Bureau.** 2002*a*. "Urban Area Criteria for Census 2000; Correction." *Federal Register: The Daily Journal of the United States*, 67: 70045–70046.

**U.S. Census Bureau.** 2002*b*. "Urban Area Criteria for Census 2000; Notice." *Federal Register: The Daily Journal of the United States*, 67: 11663–11670.

**U.S. Census Bureau.** 2011. "Urban Area Criteria for the 2010 Census; Notice." *Federal Register: The Daily Journal of the United States*, 76: 53030–53043.

**U.S. Census Bureau.** 2022. "Urban Area Criteria for the 2020 Census–Final Criteria." *Federal Register: The Daily Journal of the United States*, 87: 16706–16715.

**Van Nieuwerburgh, Stijn, and Piere-David Weill.** 2010. "Why Has Housing Dispersion Gone Up?" *The Review of Economic Studies*, 77: 1567–1606.

# Appendices and Supplemental Materials

**Table A.1: Metropolitan CBSAs Compared to KBMAs**

| metropolitan CBSA | KBMAs with core in CBSA | comparison KBMA | relative size pop | relative size emp | relative size land | share in comparison pop | share in comparison emp | share in comparison land |
|---|---|---|---|---|---|---|---|---|
| Kansas City, MO-KS | 1 | Kansas City, MO–KS | 1.23 | 1.11 | 7.33 | 0.82 | 0.90 | 0.14 |
| Los Angeles-Long Beach-Santa Ana, CA | 1 | Los Angeles-Long Beach-Santa Ana,CA | 0.81 | 0.85 | 1.21 | 0.99 | 0.99 | 0.46 |
| Riverside–San Bernardino–Ontario, CA | 2 | Los Angeles-Long Beach-Santa Ana,CA | 0.21 | 0.17 | 6.79 | 0.80 | 0.79 | 0.06 |
| Oxnard–Thousand Oaks–Ventura, CA | 2 | Oxnard, CA | 1.80 | 1.73 | 8.20 | 0.55 | 0.58 | 0.12 |
| Boston–Cambridge–Quincy, MA-NH | 4 | Boston, MA–NH–CT | 0.84 | 0.85 | 0.88 | 0.93 | 0.95 | 0.70 |
| Worcester, MA | 2 | Boston, MA–NH–CT | 0.14 | 0.12 | 0.38 | 0.89 | 0.94 | 0.59 |
| Manchester–Nashua, NH | 1 | Boston, MA–NH–CT | 0.07 | 0.07 | 0.22 | 0.87 | 0.91 | 0.40 |
| Providence-New Bedford-Fall River,RI-MA | 3 | Providence, RI–MA | 1.14 | 1.11 | 1.66 | 0.86 | 0.89 | 0.57 |
| Raleigh–Cary, NC | 1 | Raleigh–Durham, NC | 0.81 | 0.70 | 1.70 | 0.84 | 0.91 | 0.44 |
| Durham, NC | 1 | Raleigh–Durham, NC | 0.44 | 0.42 | 1.42 | 0.69 | 0.87 | 0.15 |
| minimum | 0 | | 0.02 | 0.02 | 0.04 | 0.26 | 0.19 | 0.00 |
| 10th percentile | 1 | | 0.90 | 0.88 | 1.84 | 0.50 | 0.69 | 0.03 |
| median | 1 | | 1.34 | 1.14 | 7.09 | 0.72 | 0.86 | 0.12 |
| 90th percentile | 2 | | 1.92 | 1.44 | 31.6 | 0.91 | 0.95 | 0.39 |
| maximum | 5 | | 2.75 | 2.07 | 276 | 1.00 | 1.00 | 1.00 |
| aggregate (all metropolitan CBSAs relative to all KBMAs) | | | 1.14 | 1.07 | 6.57 | 0.87 | 0.93 | 0.15 |

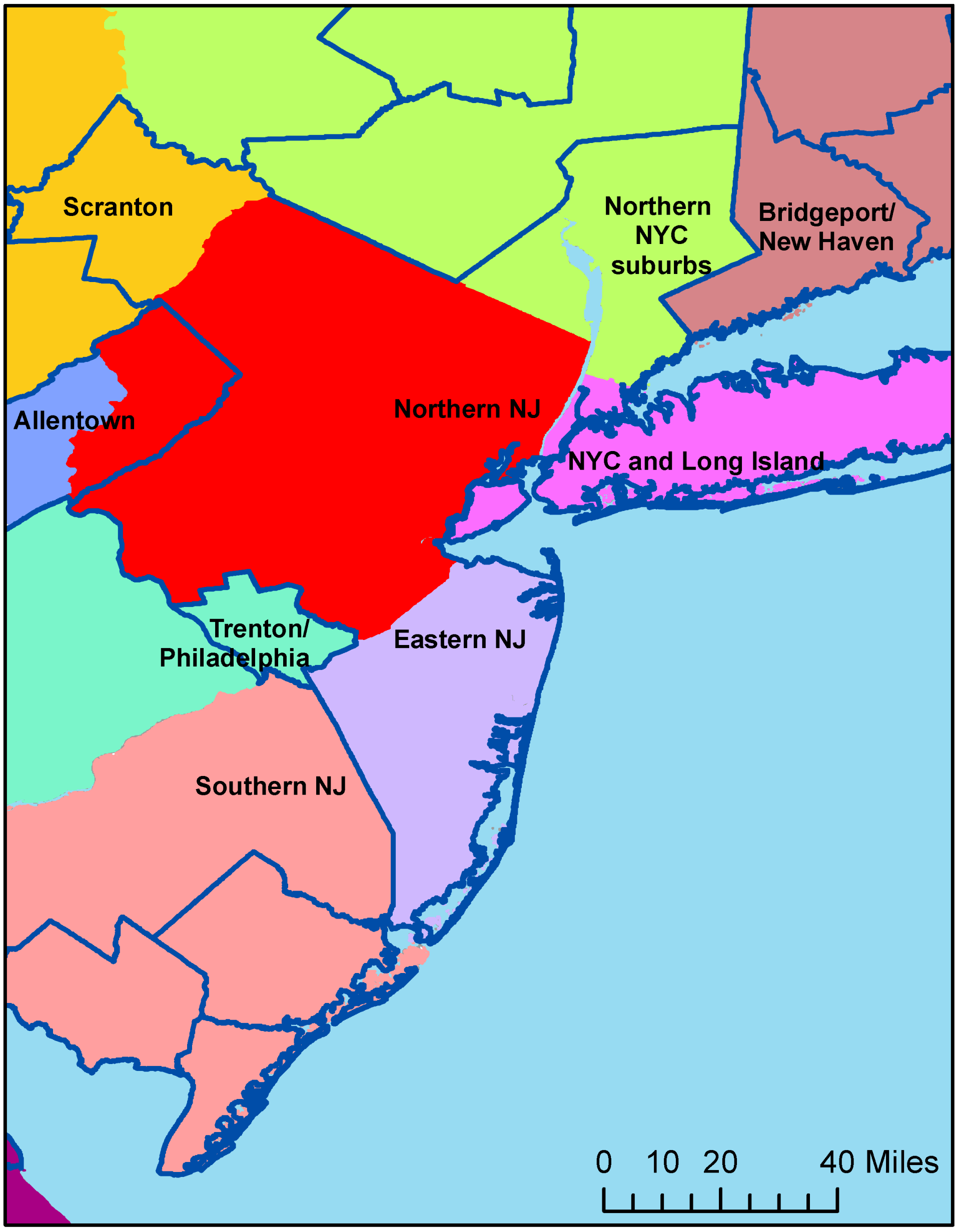
Notes: This table complements Table 4 in the main text. Each metropolitan CBSA is compared to the KBMA that has the most populous core in it. The summary statistics on relative size and overlap exclude eight metropolitan CBSAs that have no comparison KBMA, reflecting that the UA/UC cores located in them belong to built-out kernels with population below 50,000. The aggregate statistics compare all metropolitan CBSAs with all KBMAs.

**Table A.2: Monte Carlo Tests for Pareto Population Distributions**

| | Monte Carlo Simulations | | | | kernel-based metro regions | kernel-based urban areas | | |
|---|---|---|---|---|---|---|---|---|
| | F(t)= 0.050 | F(t)= 0.025 | F(t)= 0.005 | KBMAs | | | CBSAs | UAs |
| 25 obs | −7.32 | −8.77 | −11.84 | −3.27 | −5.96 | −3.16 | −2.63 | −3.47 |
| 50 obs | −9.90 | −11.86 | −15.90 | −8.82 | −10.02* | −8.94 | −9.23 | −10.27* |
| 75 obs | −11.78 | −14.00 | −19.21 | −17.09** | −17.27** | −16.31** | −16.83** | −16.64** |

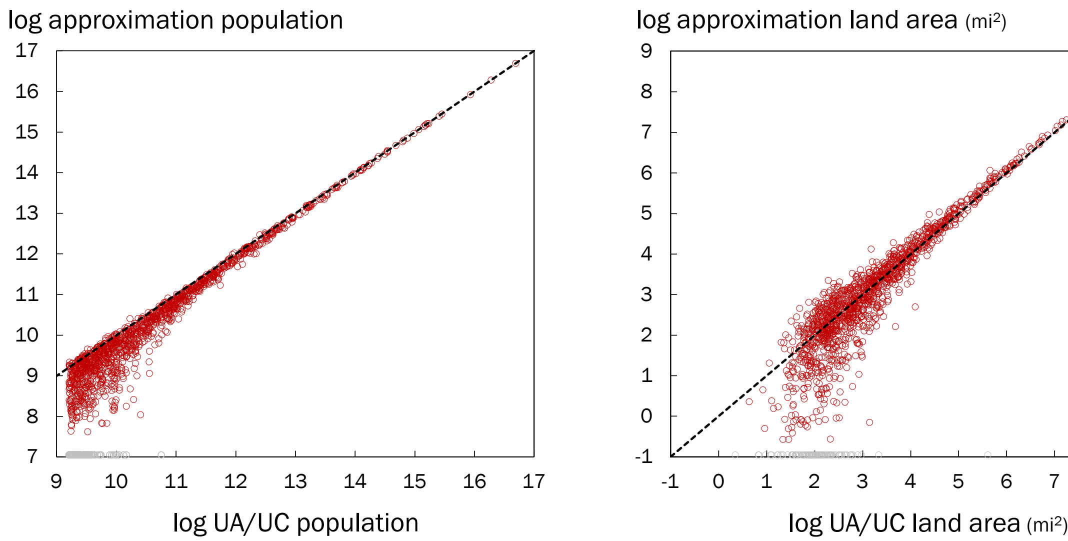
Notes: Table reports results from regressing log(rank - 0.5) on linear and quadratic log population using the largest 25, 50, and 75 observations. Subtracting 0.5 from the rank improves the fit in the presence of small-sample bias (Gabaix and Ibragimov, 2011). Following the methodology described in Gabaix and Ioannides (2004), the left three columns report the value of the t-statistic on the quadratic term at which the cumulative distribution from Monte Carlo simulations equals the specified value. The remaining columns report the t-statistic on the quadratic coefficient for each of the delineations. The Monte Carlo simulations draw the specified number of observations from a Pareto distribution 100,000 times; the t-statistic is independent of the shape parameter, reflecting that the shape parameter linearly affects both the coefficient and standard error. * denotes that we can reject that the observations are drawn from a Pareto distribution at the 0.10 level based on a two-tail test; ** that we can reject at the 0.05 level.

**Figure A.1: Commuting Zones in the Vicinity of New York City**



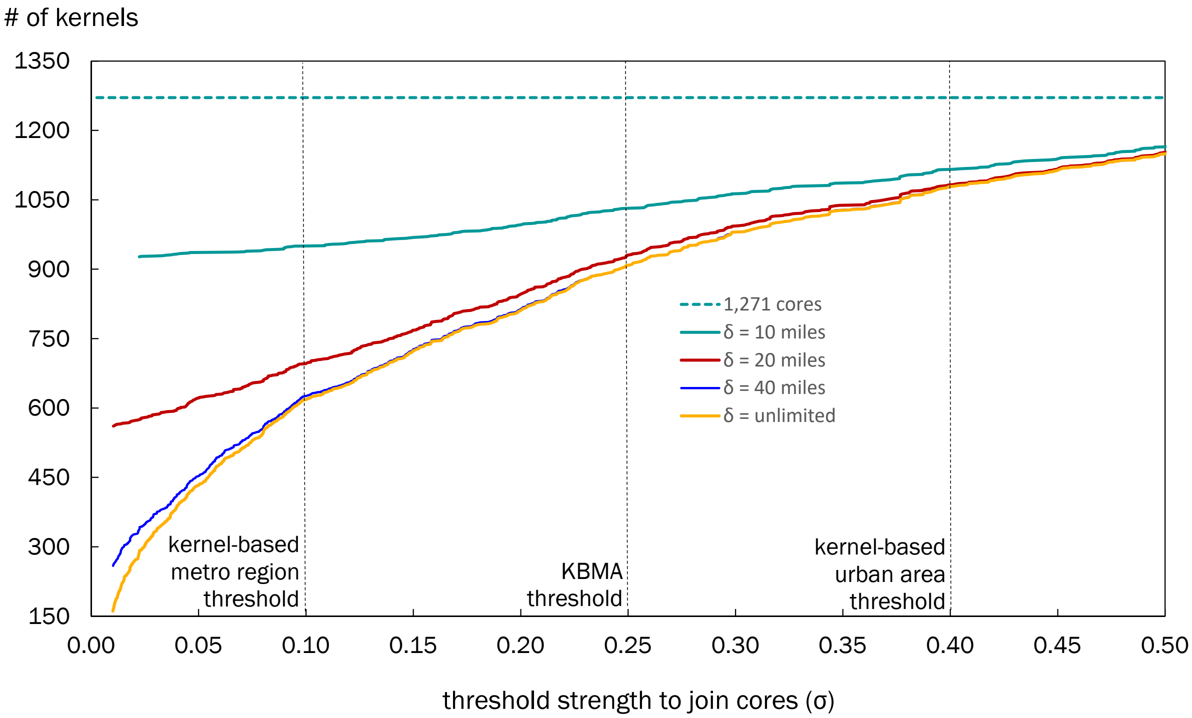

Blue lines demarcate the borders of metropolitan CBSAs. The commuting zone delineations are based on commuting patterns in 2000 (Economic Research Service, 2012). Six of the CZs are connected to Manhattan by commuter rail: NYC and Long Island, Bridgeport/New Haven, Northern NYC suburbs, Northern New Jersey, Eastern New Jersey, and Trenton/Philadelphia.

**Figure A.2: Tract-Based Approximations of UAs and UCs**



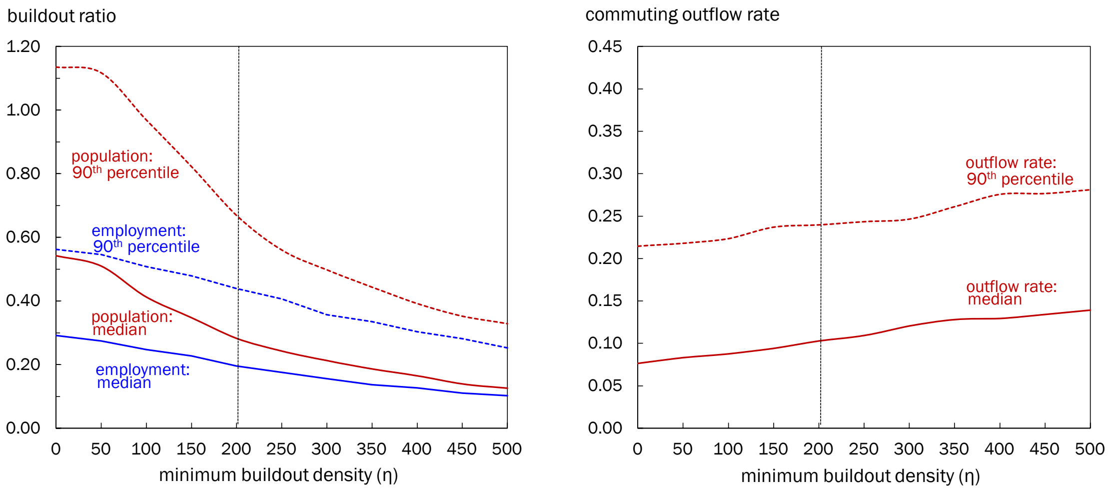

Notes: Census tracts are assigned to the approximation of a UA or a UC with population of at least 10,000 if the shares of their population and land in the UA/UC are at least 0.55 and 0.30, respectively. These thresholds minimize the sum of squared percentage differences of approximation population and land area from their actual values. Dashed lines are drawn where approximation size equals actual. Only 1,271 of the 1,372 qualifying UAs/UCs have at least one tract that meets both thresholds; the gray dots represent the 101 UCs with no tract that meets both. These null approximations only negligibly affect the delineated set of KBMAs. As illustrated by the Kansas City KBMA (Figure 4), many tracts overlapping unapproximated UCs join to a nearby kernel during the buildout. The built-out kernels of most remaining unapproximated UCs would likely fail to meet the 50,000 population threshold to qualify as a KBMA, similar to the 570 actual built-out kernels that fail to qualify.

**Figure A.3: Sensitivity of Kernel Construction to the Strength and Distance Parameters**



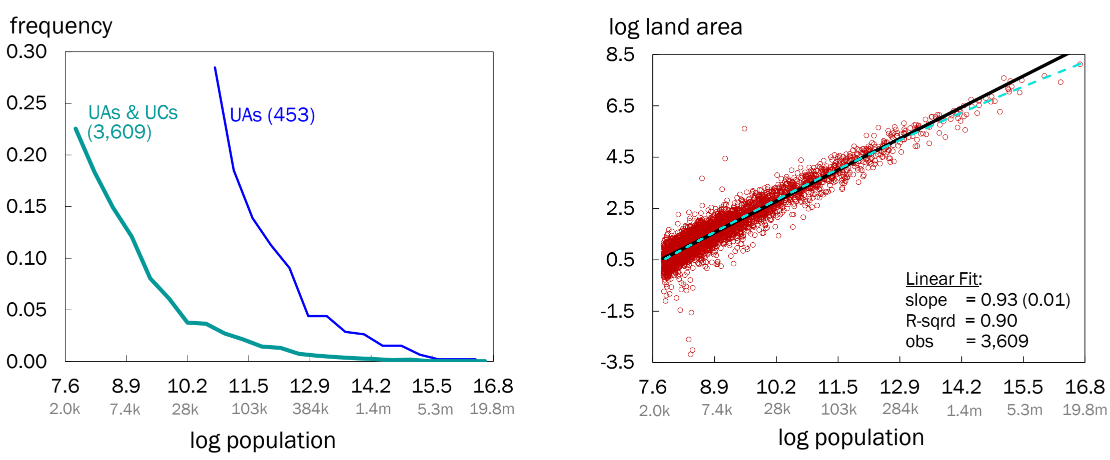

Notes: The kernel-iteration stage of the KBMA algorithm starts with 1,271 separate cores (green dashed line), corresponding to tract-based approximations of UAs and UCs with population above 10,000. Moving from right to left, the figure shows the number of surviving kernels as $\sigma$ is decreased from 0.50 to 0.01 for each of three distance maximums (green, red, and blue lines) and for unconstrained distance (yellow line). At the KBMA strength threshold, $\sigma$=0.25, the 20-mile maximum distance binds for 22 latent joins (vertical distance between the red and yellow lines). At the kernel-based urban area strength threshold, $\sigma$=0.40, the 10-mile maximum distance binds for 37 latent joins (vertical distance between the green and yellow lines). At the kernel-based metropolitan region strength threshold, $\sigma$=0.10, the 40-mile maximum distance binds for 8 latent joins (distance between the blue and yellow lines). The online supplemental materials include a workbook documenting the underlying sequences of iterative joins.

**Figure A.4: Sensitivity of Population, Employment, and Commuting Outflows to the Density Parameter**



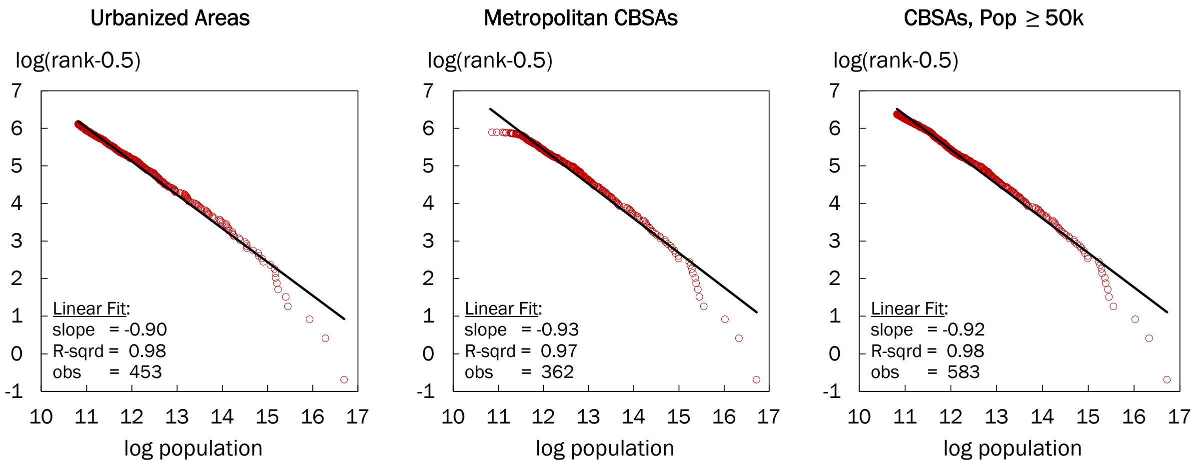

Notes: These complement the land buildout ratios and commuting inflow rates in Figure 3. Left panel shows the median and 90th-percentile ratios of the population and employment in the buildout portion of built-out kernels relative to the kernel portion as $\eta$ is increased from 0 to 500. The dashed vertical line corresponds to the parameterized KBMA value, $\eta = 200$. The vertical scale is one-tenth that used for the land buildout ratio in Figure 3, reflecting that population and employment are considerably less sensitive to $\eta$. The right panel shows the median and 90th percentile rates of commuting outflows, which are less sensitive to $\eta$ than the corresponding inflow rates shown in Figure 3. For comparability, kernels are restricted to the 302 with population of at least 50,000.

**Figure A.5: The Population and Land Area of UAs and UCs**



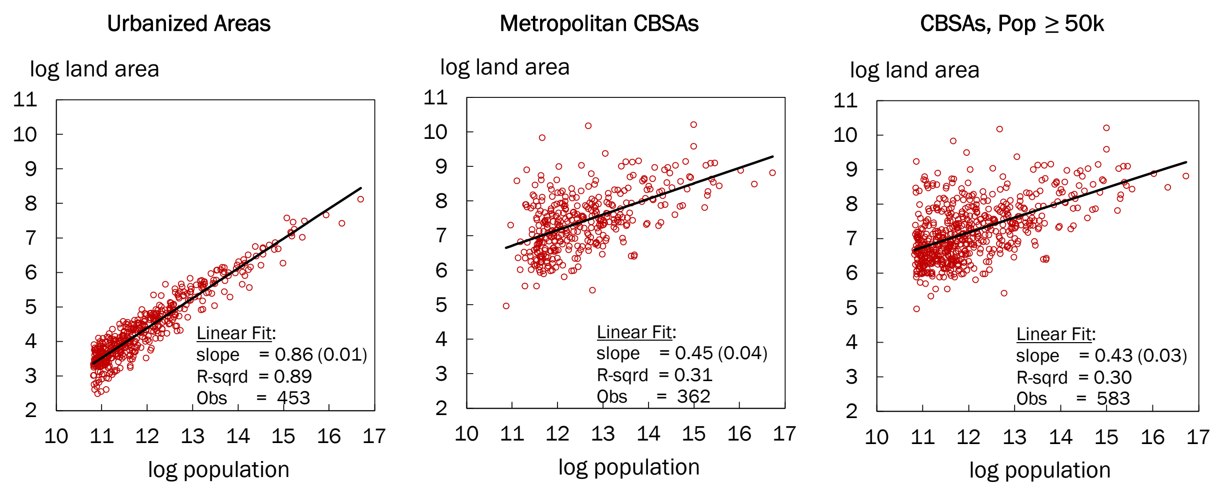

Notes: The left panel shows the population distributions of UAs and the union of UAs and UCs. Its horizontal axis measures the log of population at the lower bound of bins with width of 0.4 log points. The 3,609 UAs and UCs in 2000 together accounted for 79 percent of the U.S. population, leaving 59 million residents to be accommodated in analogous clusters with population below 2,500. The right panel plots the land area of UAs and UCs against their population. The solid and dashed lines respectively represent linear and quadratic fits. The negative estimated quadratic coefficient for the quadratic fit statistically differs from 0 at the 0.01 level. The slope of the quadratic fit, which corresponds to the implicit elasticity of land area with respect to population, declines from 1.01 at the population of the smallest UA/UC to 0.71 at the population of the largest.

## Figure A.6: Population Rank versus Size for UAs and CBSAs



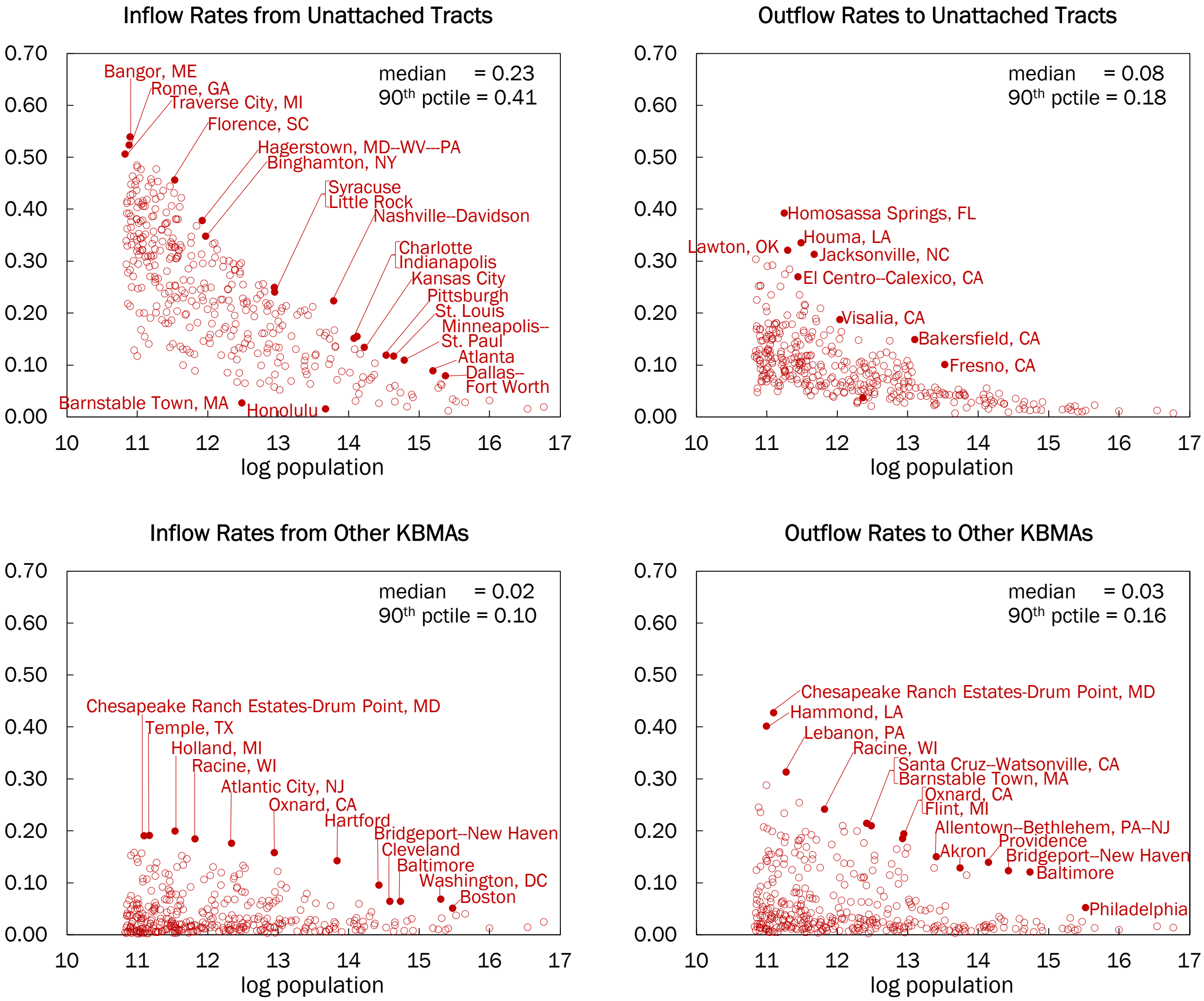

Note: Subtracting 0.5 from the rank improves the fit in the presence of small-sample bias (Gabaix and Ibragimov, 2011).

## Figure A.7: Land Area versus Population for UAs and CBSAs



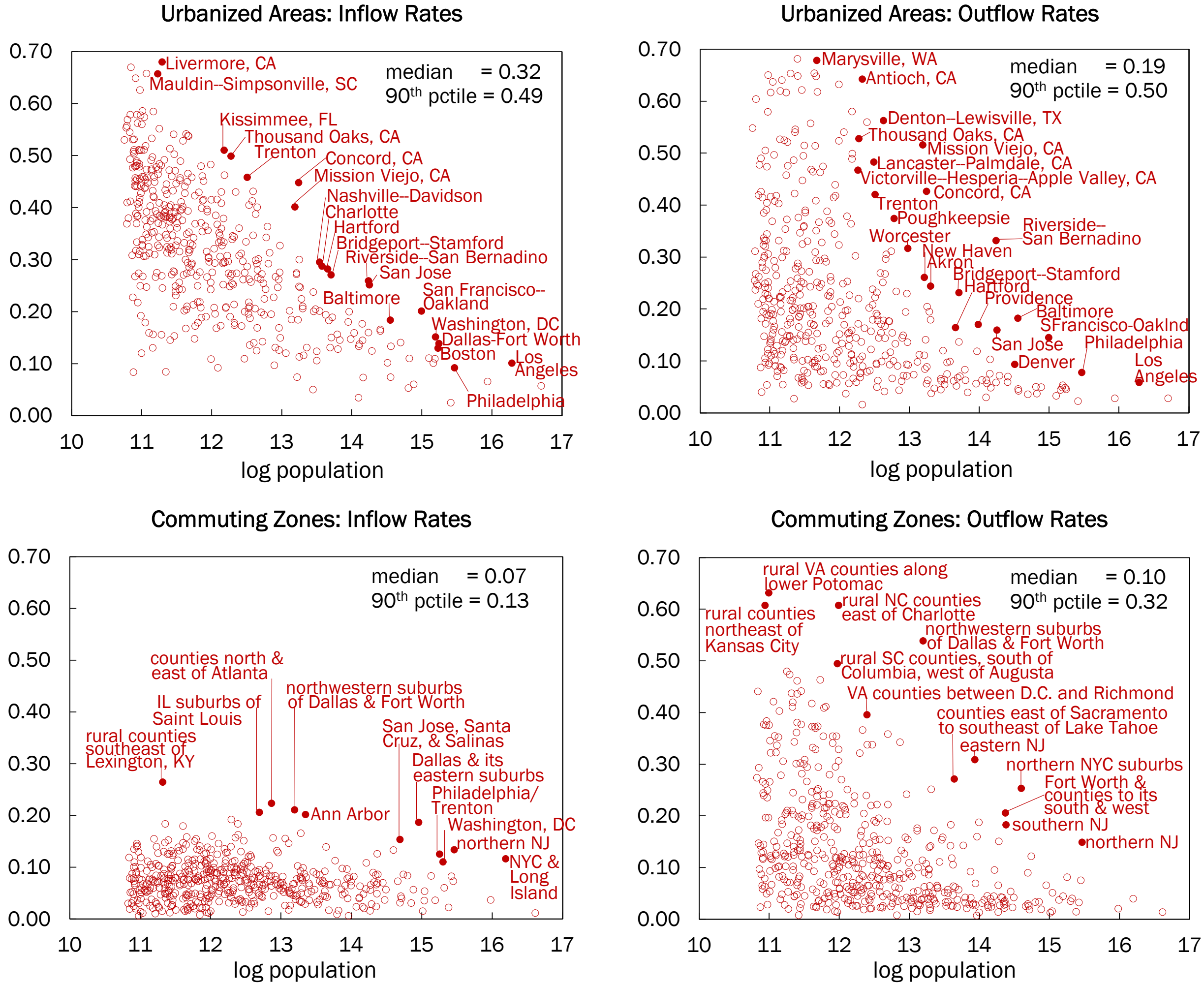

Note: Standard errors for the slope coefficient are reported in parentheses.

## Figure A.8: Decomposition of KBMA Commuting Flows




Notes: The top panels show inflow rates to KBMAs and outflow rates from them with respective origins and destinations in tracts that are not part of a KBMA. The bottom panels show inflow rates to KBMAs and outflow rates from them with respective origins and destinations in another KBMA. Commuting inflows overwhelmingly originate from unattached tracts. Commuting outflows split about equally to unattached tracts and to other KBMAs.

**Figure A.9: Commuting Flows for UAs and Commuting Zones**




Notes: Commuting inflow rates are measured as a share of employment. Commuting outflow rates are measured as a share of residents who are employed. Delineations and commuting flows are based on the 2000 decennial census. The Hightstown, NJ UA, has the highest commuting inflow and outflow rates, 0.74 and 0.81, respectively (above the displayed range). One additional UA has an inflow rate above the displayed range and nine additional ones have outflow rates above the displayed range. We use tract-based approximations of UAs to access commuting data, constructed with more relaxed threshold allocation factors than those for the kernel-based delineations. Specifically, we assign a tract to the tract-based approximation of a UA if at least 50 percent of its population is located in the UA, regardless of how much of its land area is in the UA. Doing so delineates UA approximations with population that more closely approximates the actual UA value. The associated inclusion of more unsettled land with dropping the land threshold only negligibly affects the measured commuting rates. The Commuting Zone delineations are disseminated by the Economic Research Service (Economic Research Service, 2012).